Technical Guide for Latent GOLD 5.1:

Basic, Advanced, and Syntax[1]

Jeroen K. Vermunt and Jay Magidson

Version: January 9, 2016

Statistical Innovations Inc.

(617) 489-4490

http://www.statisticalinnovations.com

[1]This document should be cited as "J.K. Vermunt and J. Magidson (2016). Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax. Belmont, MA: Statistical Innovations Inc."

For more information about Statistical Innovations Inc. please visit our website at

http://www.statisticalinnovations.com or contact us at

Statistical Innovations Inc.
375 Concord Avenue, Suite 007
Belmont, MA 02478
e-mail: support@statisticalinnovations.com

Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax.

Contents

Part I: Basic Model Options, Technical Settings, and Output Sections

1 Introduction to Part I: Basic Models

Latent GOLD 5.1 Basic implements the most important types of latent class (LC) and finite mixture (FM) models in three submodules called Cluster, DFactor, and Regression. It also contains a submodule called Step3, which can be used for step-three modeling and scoring. These are tools which are often needed *after* performing an analysis with Cluster, DFactor, or Regression.

Let us first look at the Cluster, DFactor, and Regression submodules. The differences between these arise from the fact that the application types of latent class analysis differ with respect to the required data organization and nature of the latent variables (see table).

	Data Format	
Latent Variable(s)	Single record with T responses y_{it}	T_i records with single response y_{it}
Single nominal x	Cluster	Regression
Multiple ordinal x_1, x_2,..., x_L	DFactor	–

In the Cluster and DFactor submodules, it is assumed that the data is in the standard rectangular file format, in which there is a single record for each case i. The T response variables, items, or *indicators* – which are denoted by y_{it} – appear in T columns, $1 \leq t \leq T$. Because of this *multivariate* file structure, it is straightforward to deal with response variables that are not of the same scale type; that is, we may build LC and FM models for combinations of nominal, ordinal, continuous, and count variables. In the Regression submodule, on the other hand, the T_i responses of case i appear in separate records connected by an ID variable. In other words, the data is in the form of a *repeated-measures* or *two-level* file with multiple observations of the same *dependent variable*. Compared to the Cluster and DFactor submodules, the Regression submodule is more general in that the number of responses may differ across cases (as indicated by the index i in T_i) and that various types

of constraints can be imposed on the model parameters, but also somewhat less general in that the T_i responses should be of the same scale type.

The data structure also has implications for the types of exogenous variables that can be used in the analysis. In the Cluster and DFactor submodule, one can use exogenous variables that vary between cases and that may be used to predict class membership. In Latent GOLD, such exogenous variables are called *covariates*, and denoted as z_{ir}^{cov}, $1 \leq r \leq R$, where R is the number of covariates. LC Regression models can not only contain covariates, but also exogenous variables that may vary within cases and that are used to predict the repeated measurements of the response variable of interest. These exogenous variables are referred to as *predictors* and denoted by z_{itq}^{pred}, $1 \leq q \leq Q$, where Q is the number of predictors.

As mentioned above, the nature of the latent variables also differs across the three submodules. In the Cluster and Regression submodules, it is assumed that there is a single nominal latent variable x with K categories, $1 \leq x \leq K$. The categories of this nominal latent variable are called *Clusters* or *Classes*. In the DFactor submodule, on the other hand, a model may include L ordinal (or dichotomous) latent variables called discrete factors (*DFactors*), which are denoted as x_ℓ, $1 \leq \ell \leq L$. The number of levels of the ℓth DFactor equals K_ℓ. As will be explained in more detail in Section 4, a model with L DFactors is, in fact, a restricted Cluster model with $K = \prod_{\ell=1}^{L} K_\ell$ latent classes.

The differences between the submodules described above have implications for the application types corresponding to each of the three submodules. The Cluster submodule can be used to estimate standard LC models for categorical indicators, as well as mixture-based clustering models for continuous and mixed indicators. One may include covariates to predict class membership, and because of its multivariate data structure, it is also possible to relax the local independence assumption. Cluster also implements several restricted variants of the traditional LC model, such as the LC Rasch model and the order-restricted LC model.

The LC DFactor model is a scaling tool similar to traditional factor analysis and multidimensional item response theory (IRT) models. More specifically, it can be used as an exploratory or confirmatory factor-analytic tool when not all indicators are continuous variables. The Latent GOLD program provides output similar to traditional factor analysis, such as loadings, communalities, and correlations. A one-DFactor model yields a semi-parametric variant of well-known undimensional IRT models, such as Rasch, Birnbaum,

and Nominal Response models.

The basic idea of the Regression submodule is that one specifies a regression model from the generalized linear modeling (GLM) family in which parameters differ across latent classes. One important application type is clustering or segmentation based on a regression model, for example, using data obtained by a rating-based conjoint experiment. Other applications include two-level modeling and growth modeling in which the latent class approach serves as a nonparametric random-effects modeling tool. Because of the two-level data structure and the various options for imposing parameter constraints (equality restrictions across Classes, zero and other fixed-value restrictions, and order restrictions), the Regression submodule is the most flexible submodule in terms of model constraints. For example, it can be used to estimate the most important types of restricted variants of the standard LC model, as well as multiple-group variants of the standard LC model, so long as the indicators are of the same type.

The Step3 submodule can be used for two purposes: bias-adjusted step-three modeling and computation of the scoring equation. These activities are relevant after performing a LC analysis with Cluster, DFactor, or Regression. Step-three modeling can be used to investigate the relationship between latent classes and external variables using the class definitions from a prior analysis. The external variables can either be covariates predicting the classes or outcome variables affected by the classes. Note that step-three modeling is a simpler alternative for including covariates in the model or including the outcomes as additional indicators in the model.

Computation of a scoring equation is useful if one desires to classify new observations based on a LC model built using Latent GOLD. The scoring equation can be used to obtain class predictions outside Latent GOLD, for example, in SPSS using the syntax file produced by Step3-Scoring.

The next sections describe the various types of LC and FM models implemented in Latent GOLD 5.1 Basic in more detail. First, we present the main components needed in any Latent GOLD model. Then, attention is paid to the three special cases "LC Cluster", "LC DFactor", and "LC Regression", and to step-three modeling and scoring as implemented in the Step3 submodule. Subsequently, we describe estimation procedures and the corresponding technical options. The output provided by the various submodules is described in the last section of this part of the manual.

2 Components of a Latent GOLD Model

The definition of a LC or FM model typically consists of three parts:

1. the assumed probability structure, which defines the relevant set of conditional independence assumptions among the variables in the model,
2. the assumed distributional forms for the response variables, which will depend on the scale types of the variables concerned,
3. the regression-type constraints used to gain parsimony in the description of the relationships between the variables in the model.

Latent GOLD automatically sets up the correct probability structure for each of the three special cases, Cluster, DFactor, and Regression. Before presenting these special cases, we describe the general probability structure, the available distribution functions for the indicators and dependent variables, and the corresponding regression-type constraints. More details about the exact form of these three components in the three submodules can be found in Sections 3, 4, and 5. The last subsection provides details on the coding of nominal variables appearing in the regression equations.

2.1 Probability Structure

Above we already introduced the most important notation – y_{it}, x, z_{ir}^{cov}, and z_{itq}^{pred}. It should be noted that the symbol x can also refer to the latent classes in a DFactor model, where x is obtained by combining the x_ℓ values of the multiple discrete factors. We will use bold face for vectors; that is, the symbols $\mathbf{y}_i$, $\mathbf{z}_i^{cov}$, and $\mathbf{z}_i^{pred}$ refer to the entire set of responses, covariate values, and predictor values of case i. With $\mathbf{z}_i$, we refer to all exogenous variables without making a distinction between covariates and predictors. We will also use the symbol $\mathbf{y}_{ih}$ to denote one of the H subsets of y_{it} variables, and T_h^* to denote the number of variables in subset h. By allowing a grouping of indicators into subsets, it is possible to specify models with (local) dependencies between indicator within latent classes.

Each of the Latent GOLD submodules is based on the same general mixture model probability structure that defines the relationships between the exogenous, latent, and response variables:

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{x=1}^{K} P(x|\mathbf{z}_i)\, f(\mathbf{y}_i|x,\mathbf{z}_i) = \sum_{x=1}^{K} P(x|\mathbf{z}_i) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|x,\mathbf{z}_i) \qquad (1)$$

As can be seen, we are specifying a model for $f(\mathbf{y}_i|\mathbf{z}_i)$, which is the probability density corresponding to a particular set of $\mathbf{y}_i$ values given a particular set of $\mathbf{z}_i$ values. The middle part of equation (1) shows that the unobserved variable x intervenes between the $\mathbf{z}_i$ and the $\mathbf{y}_i$ variables. Here, $P(x|\mathbf{z}_i)$ is the probability of belonging to a certain latent class given an individual's realized covariate values (the mixing weights), and $f(\mathbf{y}_i|x, \mathbf{z}_i)$ is the probability density of $\mathbf{y}_i$ given x and $\mathbf{z}_i$ (the mixture densities). Thus, x variables may be influenced by z variables, and y variables may be influenced by x and z variables.

The last part of the model formulation described in equation (1) implies that y variables belonging to different sets are assumed to be mutually independent given the latent and exogenous variables:

$$f(\mathbf{y}_i|x, \mathbf{z}_i) = \prod_{h=1}^{H} f(\mathbf{y}_{ih}|x, \mathbf{z}_i).$$

On the other hand, it is also important to note that the y's belonging to the same set h may be correlated within classes.

Let us look at some simpler special cases of the general model given in equation (1). One of these is the basic LC Cluster model that assumes local independence among all indicators and that does not include covariates; that is,

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{x=1}^{K} P(x)\, f(\mathbf{y}_i|x, \mathbf{z}_i) = \sum_{x=1}^{K} P(x) \prod_{t=1}^{T} f(y_{it}|x).$$

Each of the $f(y_{it}|x)$ is now a univariate probability density. Moreover, the mixing weights or prior class membership probabilities do not depend on covariates.

Another special case is the LC or FM Regression model. Also in these models, we assume local independence among the multiple responses y_{it}. Typical of Regression models is that the number of replications T_i may differ across cases and that there is a distinction between covariates and predictors. The probability structure then becomes

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{x=1}^{K} P(x|\mathbf{z}_i^{cov})\, f(\mathbf{y}_i|x, \mathbf{z}_{it}^{pred}) = \sum_{x=1}^{K} P(x|\mathbf{z}_i^{cov}) \prod_{t=1}^{T_i} f(y_{it}|x, \mathbf{z}_{it}^{pred}).$$

2.2 Conditional Distributions

Depending on the scale types of the variables in a set, a particular distributional form is assumed for $\mathbf{y}_{ih}$. A set may consist of one or more categorical (nominal or ordinal) variables, one or more continuous variables, or a single count variable. When the variables are categorical, a multinomial distribution is assumed for $\mathbf{y}_{ih}$. For continuous variables, we use (multivariate/censored/truncated) normal distributions. Counts can be modeled via (truncated/overdispersed) Poisson or (truncated/overdispersed) binomial distributions.[2] When referring to the distributions of discrete response variables (ordinal, nominal, Poisson count, or binomial count), we will use the symbol $P(\cdot)$ instead of $f(\cdot)$ to indicate that we are dealing with a probability instead of a density function.

Within the context of the generalized linear modeling (GLM) framework, the assumed distribution function for a response variable is denoted as the error function. Moreover, the transformation of the expected value of the response variable that yields the linear predictor that can be restricted by a regression model, is referred to as the link function (McCullagh and Nelder, 1983). Below, we define the linear predictors and the corresponding regression models for categorical, count, and continuous response variables.

2.2.1 Nominal and ordinal dependent variables

Let us first have a look at the univariate case that arises if response variable t is independent of the other response variables given x and $\mathbf{z}_i$. Let m denote a particular category of y_{it} and M_t the number of categories, where, $1 \leq m \leq M_t$. Nominal and ordinal dependent variables are assumed to come from a *multinomial* distribution with M_t entry, which means that the distribution for each of y_{it} is of the form

$$P(y_{it} = m|x, \mathbf{z}_i) = \pi_{m|t,x,\mathbf{z}_i} = \frac{\exp(\eta^t_{m|x,\mathbf{z}_i})}{\sum_{m'=1} \exp(\eta^t_{m'|x,\mathbf{z}_i})}.$$

[2] In the Regression submodule, it is also possible to use zero-inflated variants of these distributions, which amounts to adding one or more classes that give a certain response with probability one.

Here, $\pi_{m|t,x,\mathbf{z}_i}$ is the probability of giving response m given x and $\mathbf{z}_i$.[3] Furthermore, $\eta^t_{m|x,\mathbf{z}_i}$ denotes the linear term that can be further restricted by a regression model, yielding a multinomial logistic regression model with a nominal response variable and an adjacent-category ordinal logistic regression model with an ordinal response variable.

In the multivariate case, we get

$$P(\mathbf{y}_{ih} = \mathbf{m}|x, \mathbf{z}_i) = \pi_{\mathbf{m}|h,x,\mathbf{z}_i} = \frac{\exp(\eta^h_{\mathbf{m}|x,\mathbf{z}_i})}{\sum_{\mathbf{m}'} \exp(\eta^h_{\mathbf{m}'|x,\mathbf{z}_i})},$$

that is, a multinomial distribution with $M^*_h = \prod_{t\in h} M_t$ entries formed by the joint categorical variable $\mathbf{y}_{ih}$ formed by cross-tabulating the categories of the variables in subset h. The linear model for $\eta^h_{\mathbf{m}|x,\mathbf{z}_i}$ will yield a multivariate logistic model.

2.2.2 Continuous dependent variables

In the case of a single continuous y_{it}, we assume that it comes from a *univariate normal* distribution; i.e.,

$$f(y_{it}|x, \mathbf{z}_i) = \frac{1}{\sqrt{2\pi\sigma^2_{t,x}}} \exp\left\{-\frac{1}{2}\frac{\left(y_{it} - \mu_{t,x,\mathbf{z}_i}\right)^2}{\sigma^2_{t,x}}\right\}, \tag{2}$$

where $\mu_{t,x,\mathbf{z}_i}$ and $\sigma^2_{t,x}$ are the mean and the variance, respectively.

In the multivariate case, continuous dependent variables are modeled by means of *multivariate normal* distributions; that is,

$$\begin{aligned} f(\mathbf{y}_{ih}|x, \mathbf{z}_i) &= (2\pi)^{-K_h/2} \left|\mathbf{\Sigma}_{h,x}\right|^{-1/2} \\ &\quad \exp\left\{-\frac{1}{2}\left(\mathbf{y}_{ih} - \boldsymbol{\mu}_{h,x,\mathbf{z}_i}\right)' \mathbf{\Sigma}^{-1}_{h|x} \left(\mathbf{y}_{ih} - \boldsymbol{\mu}_{h,x,\mathbf{z}_i}\right)\right\}. \end{aligned} \tag{3}$$

Here, the vector $\boldsymbol{\mu}_{h,x,\mathbf{z}_i}$ contains the conditional expectations of the y variables belonging to set h, and $\mathbf{\Sigma}_{h,x}$ their variances and covariances. As is indicated by their indices, expectations may depend both on the latent class to which one belongs and on an individual's covariate values, whereas variances and covariances may only be class dependent.

[3] The multinomial probability density for a single "trial" is sometimes written as $\prod_{m=1}^{M_t}(\pi_{m|t,x,\mathbf{z}_i})^{\delta_{itm}}$, where δ_{itm} is an indicator variable taking the value 1 if $y_{it} = m$, and otherwise 0.

Two univariate variants that can be used for continuous dependent variables are left truncated and left censored normal distributions. The *truncated normal* model can be used if only cases with $y_{it} > 0$ are included in the sample. Its form is

$$f(y_{it}|x, \mathbf{z}_i, y_{it} > 0) = \frac{f(y_{it}|x, \mathbf{z}_i)}{1 - F(0|\mu_{t,x,\mathbf{z}_i}, \sigma^2_{t,x})},$$

where $f(y_{it}|x, \mathbf{z}_i)$ is the univariate normal density defined in equation (2) and $F(0|\mu_{t,x,\mathbf{z}_i}, \sigma^2_{t,x})$ the cumulative univariate normal distribution evaluated at $y = 0$.

The *censored normal* model is useful if $y_{it} \geq 0$ for all i and t, but with many more $y_{it} = 0$ than can be expected based on a univariate normal distribution. A censored normal distribution is obtained by assuming a univariate normal distribution $f(y_{it}|x, \mathbf{z}_i)$ if $y_{it} > 0$ and a cumulative univariate normal distribution $F(0|\mu_{t,x,\mathbf{z}_i}, \sigma^2_{t,x})$ if $y_{it} = 0$.

In each of the normal models, the linear predictor is simply the expectation of the distribution: $\mu_{t,x,\mathbf{z}_i} = \eta^t_{x,\mathbf{z}_i}$. Restricting the linear term $\eta^t_{x,\mathbf{z}_i}$ can be achieved by a standard linear regression model.

2.2.3 Poisson counts

A count or a number of events can be modeled by a *Poisson* distribution. The form of this distribution is

$$\begin{aligned} P(y_{it}|x, \mathbf{z}_i, e_{it}) &= \frac{1}{y_{it}!}(\theta_{t,x,\mathbf{z}_i} e_{it})^{y_{it}} \exp(-\theta_{t,x,\mathbf{z}_i} e_{it}) \\ &= \frac{1}{y_{it}!}(\mu_{t,x,i})^{y_{it}} \exp(-\mu_{t,x,i}). \end{aligned}$$

Here, $\theta_{t,x,\mathbf{z}_i}$ denotes the Poisson rate and e_{it} the exposure of case i to event t, and $\mu_{t,x,i} = \theta_{t,x,\mathbf{z}_i} e_{it}$. In the Cluster and DFactor submodules, the exposure is assumed to be equal to 1 for all cases, whereas in the Regression submodule the exposure is a variable that can be specified by the user.

The *truncated Poisson* distribution, which can be used if only cases with nonzero counts are included in the sample, has the form

$$P(y_{it}|x, \mathbf{z}_i, e_{it}, y_{it} > 0) = \frac{P(y_{it}|x, \mathbf{z}_i, e_{it})}{1 - P(0|x, \mathbf{z}_i, e_{it})}, \tag{4}$$

where $P(y_{it}|x, \mathbf{z}_i, e_{it})$ is the Poisson probability of having y_{it} events and $P(0|x, \mathbf{z}_i, e_{it}) = \exp(-t, x, i)$ is the Poisson probability of having zero events.

In the Regression submodule, it is also possible to specify models for overdispersed Poisson counts. By assuming that the Poisson rate follows a gamma distribution, one obtains a *negative-binomial distribution* (Long, 1997; Simonoff, 2003). The exact form of this distribution is

$$P(y_{it}|x, \mathbf{z}_i, e_{it}) = \frac{\Gamma(y_{it} + 1/\sigma^2_{t,x})}{y_{it}!\,\Gamma(1/\sigma^2_{t,x})} \left(\frac{1/\sigma^2_{t,x}}{1/\sigma^2_{t,x} + \mu_{t,x,i}}\right)^{1/\sigma^2_{t,x}} \left(\frac{\mu_{t,x,i}}{1/\sigma^2_{t,x} + \mu_{t,x,i}}\right)^{y_{it}},$$

where $\Gamma(\cdot)$ is the gamma function. The additional parameter $\sigma^2_{t,x}$ denotes the variance of the gamma mixing distribution, also referred to as the dispersion parameter.[4] As in the standard Poisson case, the class-specific expected value of y_{it} equals $\mu_{t,x,i}$. The variance is, however, no longer equal to the expected value $\mu_{t,x,i}$, but is a factor $1 + \sigma^2_{t,x}\,\mu_{t,x,i}$ larger. In other words, the larger the value of $\sigma^2_{t,x}$, the larger the amount of overdispersion. Note that for $\sigma^2_{t,x} = 0$ we have again a Poisson distribution.

A *truncated negative-binomial* distribution is defined in the same manner as a truncated Poisson distribution (see equation 4), with the difference that now $P(y_{it}|x, \mathbf{z}_i, e_{it})$ and $P(0|x, \mathbf{z}_i, e_{it})$ are negative-binomial probabilities of having y_{it} and zero events.

The Poisson rate $\theta_{t,x,\mathbf{z}_i}$ can be written in terms of a linear predictor $\eta^t_{x,\mathbf{z}_i}$ as follows:

$$\theta_{t,x,\mathbf{z}_i} = \exp(\eta^t_{x,\mathbf{z}_i}),$$

or

$$\mu_{t,x,i} = \exp(\eta^t_{x,\mathbf{z}_i})\, e_{it},$$

which yields the well-known log-linear Poisson regression model.

2.2.4 Binomial counts

Counts can also be modeled by a binomial model for multiple trials. The *binomial* distribution for counts equals

$$P(y_{it}|x, \mathbf{z}_i, e_{it}) = \left(\frac{e_{it}!}{y_{it}!\ (e_{it} - y_{it})!}\right) (\pi_{t,x,\mathbf{z}_i})^{y_{it}} (1 - \pi_{t,x,\mathbf{z}_i})^{(e_{it} - y_{it})}.$$

[4] Alternatives notations are α for the dispersion parameter and ν for its inverse.

Here, e_{it} denotes the total number of trials, which corresponds to the maximum number of events (or successes) that an individual can experience at replication t. In the Cluster and DFactor submodules, the total is assumed to be equal to the largest observed y_{it} in the sample. In the Regression submodule, on the other hand, the total (exposure) is a variable that can be specified by the user.

A *truncated binomial* distribution is defined in the same manner as a truncated Poisson distribution (see equation 4), with the difference that now $P(y_{it}|x, \mathbf{z}_i, e_{it})$ and $P(0|x, \mathbf{z}_i, e_{it}) = (1 - \pi_{t,x,\mathbf{z}_i})^{e_{it}}$ are binomial probabilities of having y_{it} and zero events.

In the Regression submodule, it is also possible to specify models for overdispersed binomial counts. By assuming that the success probabilities follow a beta distribution, one obtains a *beta-binomial distribution* (Agresti, 2000; Simonoff, 2003). The exact form of this distribution is

$$P(y_{it}|x, \mathbf{z}_i, e_{it}) = \frac{B\left(\pi_{t,x,\mathbf{z}_i}/\sigma^2_{t,x} + y_{it}, (1 - \pi_{t,x,\mathbf{z}_i})/\sigma^2_{t,x} + (e_{it} - y_{it})\right)}{B\left(\pi_{t,x,\mathbf{z}_i}/\sigma^2_{t,x}, (1 - \pi_{t,x,\mathbf{z}_i})/\sigma^2_{t,x}\right)},$$

where $B(\cdot)$ is the beta function. The parameter $\sigma^2_{t,x}$, which is referred to as the dispersion parameter, affects the variance of the beta mixing distribution.[5] As in the standard binomial case, the class-specific expected value of y_{it} equals $\pi_{t,x,\mathbf{z}_i} e_{it}$. The variance is, however, no longer equal to $\pi_{t,x,\mathbf{z}_i}(1 - \pi_{t,x,\mathbf{z}_i})e_{it}$, but is a factor $1 + (e_{it} - 1)\sigma^2_{t,x}/(1 + \sigma^2_{t,x})$ larger. In other words, the larger the value of $\sigma^2_{t,x}$, the larger the amount of overdispersion. For $\sigma^2_{t,x} = 0$, the beta-binomial distribution reduces to the standard binomial distribution.

A *truncated beta-binomial* distribution is defined in the same manner as a truncated Poisson distribution (see equation 4), with the difference that now $P(y_{it}|x, \mathbf{z}_i, e_{it})$ and $P(0|x, \mathbf{z}_i, e_{it})$ are beta-binomial probabilities of having y_{it} and zero events.

The parameterization of the binomial probability $\pi_{t,x,\mathbf{z}_i}$ is similar to the one for dichotomous categorical y variables; i.e.,

$$\pi_{t,x,\mathbf{z}_i} = \frac{\exp(\eta^t_{x,\mathbf{z}_i})}{1 + \exp(\eta^t_{x,\mathbf{z}_i})}.$$

[5] An alternative formulation uses the symbol α for the term π/σ^2 and β for $(1-\pi)/\sigma^2$, with $\pi = \alpha/(\alpha + \beta)$ and $\sigma^2 = 1/(\alpha + \beta)$.

As can be seen, this yields a binary logistic regression model with linear predictor $\eta^t_{x,\mathbf{z}_i}$.

2.2.5 Latent variables

As nominal and ordinal dependent variables, the values on the latent variables given a person's covariate values are assumed to come from a (joint) multinomial distribution. The *multinomial* probability $P(x|\mathbf{z}_i)$ is parameterized as follows:

$$P(x|\mathbf{z}_i) = \pi_{x|\mathbf{z}_i} = \frac{\exp(\eta_{x|\mathbf{z}_i})}{\sum_{x'=1}^{K} \exp(\eta_{x'|\mathbf{z}_i})}.$$

In the Cluster and Regression submodules, we have a single nominal latent variable, which means that this yields a standard multinomial logit model. In DFactors models, depending on whether we have a single or multiple ordinal latent variables, we have a standard or a multivariate version of the adjacent-category ordinal logit model.

2.3 Types of GLM-family Regression Models

The exact form that the four linear predictors introduced above – $\eta^t_{x,\mathbf{z}_i}$, $\eta^t_{m|x,\mathbf{z}_i}$, $\eta^h_{\mathbf{m}|x,\mathbf{z}_i}$, and $\eta_{x|\mathbf{z}_i}$ – take on in the Cluster, DFactor, and Regression submodules discussed is given later on. Here, we will provide a more generic formulation of the regression models of interest, in which we use the symbol y for the outcome variable that could also be a latent variable (in $\eta_{x|\mathbf{z}_i}$) and z_{ip} for an explanatory variable that could also be a latent variable (in $\eta^t_{x,\mathbf{z}_i}$, $\eta^t_{m|x,\mathbf{z}_i}$, and $\eta^t_{\mathbf{m}|x,\mathbf{z}_i}$).

With a *continuous response variable*, we have a standard linear regression model

$$\eta_{\mathbf{z}_i} = \mu_{\mathbf{z}_i} = \beta_0 + \sum_{p=1}^{P} \beta_p \cdot z_{ip},$$

with intercept β_0 and regression coefficients β_p.

For *Poisson counts*, we use a log-linear Poisson model and for *binomial counts* a binary logistic regression model:

$$\eta_{\mathbf{z}_i} \quad = \quad \log(\theta_{\mathbf{z}_i}) = \beta_0 + \sum_{p=1}^{P} \beta_p \cdot z_{ip},$$

$$\eta_{\mathbf{z}_i} = \log\left(\frac{\pi_{\mathbf{z}_i}}{1-\pi_{\mathbf{z}_i}}\right) = \beta_0 + \sum_{p=1}^{P} \beta_p \cdot z_{ip},$$

where again β_0 and β_p are the intercept and the regression coefficients.

Slightly more complicated is the case of a *nominal outcome variable*. When using *dummy coding* with the reference category denoted by m', we get the following definition of the linear term $\eta_{m|\mathbf{z}_i}$:

$$\eta_{m|\mathbf{z}_i} = \log\left(\frac{P(y=m|\mathbf{z}_i)}{P(y=m'|\mathbf{z}_i)}\right) = \beta_{m0} + \sum_{p=1}^{P} \beta_{mp} \cdot z_{ip},$$

where $\beta_{m'p} = 0$, for $0 \leq p \leq P$. This yields what is usually referred to as the baseline-category logit model (Agresti, 2002). Note that this multinominal logistic regression model contains $M-1$ sets of free β parameters.

An alternative to dummy coding is to use *effect coding* for the dependent variable, which yields the following definition of $\eta_{m|\mathbf{z}_i}$:

$$\eta_{m|\mathbf{z}_i} = \log\left(\frac{P(y=m|\mathbf{z}_i)}{\left[\prod_{m'=1}^{M} P(y=m'|\mathbf{z}_i)\right]^{1/M}}\right) = \beta_{m0} + \sum_{p=1}^{P} \beta_{mp} \cdot z_{ip}$$

where $\sum_{m=1}^{M} \beta_{m'p} = 0$, for $0 \leq p \leq P$. As can be see, probability of answering in category m is now compared with the average (geometric mean) of the probabilities of all M categories.

With an *ordinal dependent variable* we make use of the adjacent-category ordinal logit model (Agresti, 2002; Goodman, 1979; Magidson, 1996) in which

$$\eta_{m|\mathbf{z}_i} = \beta_{m0} + \sum_{p=1}^{P} \beta_{\cdot p} \cdot y_m^* \cdot z_{ip},$$

irrespective of whether one uses dummy or effect coding for the dependent variable. As can be seen, compared to the nominal logit model, this ordinal logit model is obtained by the restriction that $\beta_{mp} = \beta_p \cdot y_m^*$, where y_m^* is the score assigned to category m of the dependent variable.[6] What one will typically interpret are the $M-1$ adjacent-category logits

$$\log\left(\frac{P(y=m+1|\mathbf{z}_i)}{P(y=m|\mathbf{z}_i)}\right) = \eta_{m+1|\mathbf{z}_i} - \eta_{m|\mathbf{z}_i} = \beta_{m0}^* + \sum_{p=1}^{P} \beta_{\cdot p} \cdot (y_{m+1}^* - y_m^*) \cdot z_{ip},$$

[6] In the case that the y_m^* scores sum to 0, the linear predictor can be interpreted as with effect coding of a nominal dependent variable. When the score for the first category, y_1^*, is set equal to 0, one gets a baseline-category logit, with $m = 1$ as reference category.

where $\beta^*_{m0} = \beta_{m+1,0} - \beta_{m0}$. In the case of category scores with an equal mutual distance of 1, the adjacent-category logits simplify to

$$\log\left(\frac{P(y = m+1|\mathbf{z}_i)}{P(y = m|\mathbf{z}_i)}\right) = \beta^*_{m0} + \sum_{p=1}^{P} \beta_{\cdot p} \cdot z_{ip}.$$

As can be seen, in fact, we have a standard binary logistic regression model for each pair of adjacent categories, where the intercept is pair specific but the regression weights are equal across pairs.

Let us now look at the case of a *multivariate nominal/ordinal response variable*. Assume that the set of interest consists of two nominal variables y_1 and y_2. In that case, the linear predictor in the model for the joint multinomial probability will be of the form

$$\eta_{m_1,m_2|\mathbf{z}_i} = \beta^1_{m_1 0} + \sum_{p=1}^{P} \beta^1_{m_1 p} \cdot z_{ip} + \beta^2_{m_2 0} + \sum_{p=1}^{P} \beta^2_{m_2 p} \cdot z_{ip} + \beta^{12}_{m_1 m_2},$$

where the superscripts of the β parameters indicate to which dependent variable the term concerned belongs. As can be seen, the first two terms correspond to y_1 and the third and fourth to y_2. The new term compared to the multinomial logistic regression described above is the parameter capturing the within-class association between y_1 and y_2, denoted by $\beta^{12}_{m_1 m_2}$. Inclusion of such a two-way association term is what happens if two categorical (nominal or ordinal) variables are assumed to be locally dependent.

If both y_1 and y_2 are ordinal variables, we restrict $\beta^1_{m_1 p} = \beta^1_{\cdot p} \cdot y^*_{m_1}$, $\beta^2_{m_2 p} = \beta^2_{\cdot p} \cdot y^*_{m_2}$, and $\beta^{12}_{m_1 m_2} = \beta^{12}_{\cdot\cdot} \cdot y^*_{m_1} \cdot y^*_{m_2}$. If y_1 is nominal and y_2 ordinal, we get $\beta^2_{m_2 p} = \beta^2_{\cdot p} \cdot y^*_{m_2}$, and $\beta^{12}_{m_1 m_2} = \beta^{12}_{m_1 \cdot} \cdot y^*_{m_2}$. In other words, as in the standard adjacent-category ordinal logit model, the information that a categorical response variable is ordinal implies that its category scores are used to restrict its relationships with other variables.

2.4 Coding of Nominal Variables

In the description of the various regression models, we assumed that predictors and covariates were numeric. There is no such limitation however, as Latent GOLD allows one or more of these explanatory variables to be specified to be nominal. For nominal variables, Latent GOLD sets up the design vectors using either effect (ANOVA-type) coding or dummy coding with the

first or last category as reference category for identification. Effect coding means that the parameters will sum to zero over the categories of the nominal variable concerned, In dummy coding, the parameters corresponding to the reference category are fixed to zero.

Suppose we have a nominal predictor variable with 4 categories. The effect coding constraint implies that the corresponding 4 effects should sum to 0. This is accomplished by defining a design matrix with 3 "numeric" predictors z_{i1}, z_{i2}, and z_{i3}. The design matrix that is set up for the 3 non-redundant terms (β_1, β_2, β_3) is as follows:

$$\begin{array}{lccc} \text{category 1} & 1 & 0 & 0 \\ \text{category 2} & 0 & 1 & 0 \\ \text{category 3} & 0 & 0 & 1 \\ \text{category 4} & -1 & -1 & -1 \end{array} ,$$

where each row corresponds to a category of the explanatory variable concerned and each column to one of the three parameters. Although the parameter for the last category is omitted from the model, you do not notice that because it is computed by the program after the model is estimated. The parameter for the fourth category equals $-\sum_{p=1}^{3}\beta_p$; that is, minus the sum of the parameters of the three other categories. This guarantees that the parameters sum to zero since $\sum_{p=1}^{3}\beta_p - \sum_{p=1}^{3}\beta_p = 0$.

Instead of using effect coding, it is also possible to use dummy coding. Depending on whether one uses the first or the last category as reference category, the design matrix will look like this

$$\begin{array}{lccc} \text{category 1} & 0 & 0 & 0 \\ \text{category 2} & 1 & 0 & 0 \\ \text{category 3} & 0 & 1 & 0 \\ \text{category 4} & 0 & 0 & 1 \end{array} ,$$

or this

$$\begin{array}{lccc} \text{category 1} & 1 & 0 & 0 \\ \text{category 2} & 0 & 1 & 0 \\ \text{category 3} & 0 & 0 & 1 \\ \text{category 4} & 0 & 0 & 0 \end{array} .$$

Whereas in effect coding the category-specific effects should be interpreted in terms of deviation from the average, in dummy coding their interpretation is in terms of difference from the reference category. Note that the parameter for the reference category is omitted, which implies that it is equated to 0.

2.5 Known-Class Indicator

Sometimes, one has a priori information – for instance, from an external source – on the class membership of some individuals. For example, in a four-class situation, one may know that case 5 belongs to latent class 2 and case 11 to latent class 3. Similarly, one may have a priori information on which class cases do not belong to. For example, again in a four-class situation, one may know that case 19 does not belong to latent class 2 and that case 41 does not belong to latent classes 3 or 4. In Latent GOLD, there is an option – called "*Known Class*" – for indicating to which latent classes cases do *not* belong to.

Let $\boldsymbol{\tau}_i$ be a vector of 0-1 variables containing the "*Known Class*" information for case i, where $\tau_{ix} = 0$ if it is known that case i does not belong to class x, and $\tau_{ix} = 1$ otherwise. The vector $\boldsymbol{\tau}_i$ modifies the general probability structure defined in equation (1) as follows:

$$f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\tau}_i) = \sum_{x=1}^{K} \tau_{ix}\, P(x|\mathbf{z}_i)\, f(\mathbf{y}_i|x, \mathbf{z}_i)\,.$$

As a result of this modification, the posterior probability of belonging to class x will be equal to 0 if $\tau_{ix} = 0$.

The known-class option has three important applications.

1. It can be used to estimate models with training cases; that is, cases for which class membership has been determined using a gold standard method. Depending on how this training information is obtained, the missing data mechanism will be MCAR (Missing Completely At Random, where the known-class group is a random sample from all cases), MAR (Missing At Random, where the known-class group is a random sample given observed responses and covariate values), or NMAR (Not Missing At Random, where the known-class group is a non-random sample and thus may depend on class membership itself). MAR occurs, for example, in clinical applications in which cases with more than a certain number of symptoms are subjected to further examination to obtain a perfect classification (diagnosis). NMAR may, for example, occur if training cases that do not belong to the original sample under investigation are added to the data file.

   Both in the MAR and MCAR situation, parameter estimates will be unbiased. In the NMAR situation, however, unbiased estimation requires

that separate class sizes are estimated for training and non-training cases (McLachlan and Peel, 2000). This can easily be accomplished by expanding the model of interest with a dichotomous covariate that takes on the value 0 for training cases and 1 for non-training cases.

2. Another application is specifying models with a partially missing discrete variable that affects one or more response variables. An important example is the *complier average causal effect* (CACE) model proposed by Imbens and Rubin (1997), which can be used to determine the effect of a treatment conditional on compliance with the treatment. Compliance is, however, only observed in the treatment group, and is missing in the control group. In Latent GOLD, this CACE model can be specified as a LC Regression model, in which class membership (compliance) is known for the treatment group, and which a treatment effect is specified only for the compliance class.

3. The known-class indicator can also be used to specify *multiple-group LC models*. Suppose we have a three-class model and two groups, say males and females. A multiple-group LC model is obtained by indicating that there are six latent classes, were males may belong to classes 1–3 and females to classes 4–6. To get the correct output, the grouping variable should not only be used as the known-class indicator, but also as a nominal covariate.

3 Latent Class Cluster Models

The LC Cluster model implemented in Latent GOLD is a model with:

1. a single nominal latent variable x,

2. T response variables y_{it} (indicators) that can be nominal, ordinal, continuous, and/or counts,

3. R numeric or nominal covariates z_{ir}^{cov} affecting x,

4. direct relationships between indicators and/or direct effects of covariates on indicators.

Below we first describe the basic components of a LC Cluster model; that is, the underlying probability structure and the form of the linear terms. Then, we pay attention to the most important special cases of the Cluster model: standard LC analysis and mixture-model clustering. First, we describe the standard LC model for nominal and ordinal categorical y variables without covariates. Then, we show how covariates are included in the model. Next, we indicate how the assumption of local independence is relaxed or, in other words, how z–y and y–y relationships are included in a model. The next two parts of this section deal with LC Cluster models for continuous y variables and LC Cluster models for mixed-mode data, respectively. The last section gives an overview of possible parameter restrictions.

3.1 Probability Structure and Linear Predictors

Assuming that the model of interest contains covariates, the following structure serves as the starting point of a LC Cluster analysis:[7]

$$f(\mathbf{y}_i|\mathbf{z}_i^{cov}) = \sum_{x=1}^{K} P(x|\mathbf{z}_i^{cov}) \prod_{t=1}^{T} f(y_{it}|x). \tag{5}$$

In other words, covariates affect the latent variable (the Clusters) but have no direct effects on the indicators, and indicators are assumed to be mutually independent given cluster membership. The most general probability structure that can be used, however, also allows the inclusion of direct effects of covariates on indicators and associations/correlations between indicators within Clusters. For the latter, we have to group the T indicators into H sets as was already explained in Section 2, where the indicators belonging to the same set may be correlated within classes. The most general LC Cluster probability structure is

$$f(\mathbf{y}_i|\mathbf{z}_i^{cov}) = \sum_{x=1}^{K} P(x|\mathbf{z}_i^{cov}) \prod_{h=1}^{H} f(\mathbf{y}_{ih}|x, \mathbf{z}_i^{cov}), \tag{6}$$

where the exact form of each of the class-specific conditional distributions $f(\mathbf{y}_{ih}|x, \mathbf{z}_i^{cov})$ depends on the scale types of the variables in subset h.

[7] Without covariates, this simplies to $f(\mathbf{y}_i) = \sum_{x=1}^{K} P(x) \prod_{t=1}^{T} f(y_{it}|x)$.

In Cluster models, one can make use of four linear predictors: $\eta^t_{x,\mathbf{z}_i}$, $\eta^t_{m|x,\mathbf{z}_i}$, $\eta^h_{\mathbf{m}|x,\mathbf{z}_i}$, and $\eta_{x|\mathbf{z}_i}$. For continuous and count indicators, we get

$$\eta^t_{x,\mathbf{z}_i} = \beta^t_0 + \beta^t_{x0} + \sum_{r=1}^{R} \beta^t_r \cdot z^{cov}_{ir},$$

where β^t_0 is the intercept, β^t_{x0} the effect of the Clusters on y_{it}, and β^t_r the direct effect of covariate r on the indicator concerned. For identification, either effect or dummy coding constraints have to be imposed on β^t_{x0}; that is, either $\sum_{x=1}^{K} \beta^t_{x0} = 0$, $\beta^t_{10} = 0$, or $\beta^t_{K0} = 0$.

For nominal indicators, we use multinomial logit models having linear predictors of the form

$$\eta^t_{m|x,\mathbf{z}_i} = \beta^t_{m0} + \beta^t_{mx0} + \sum_{r=1}^{R} \beta^t_{mr} \cdot z^{cov}_{ir},$$

with restrictions $\sum_{x=1}^{K} \beta^t_{mx0} = 0$, $\beta^t_{m10} = 0$, or $\beta^t_{mK0} = 0$ for the latent variable x, and a similar set of identification constraints for the dependent variable: $\sum_{m=1}^{M_t} \beta^t_{mx0} = 0$, $\beta^t_{1x0} = 0$, or $\beta^t_{M_t x0} = 0$.

An ordinal response variable is modeled with an adjacent-category logit model in which

$$\eta^t_{m|x,\mathbf{z}_i} = \beta^t_{m0} + \beta^t_{\cdot x0} \cdot y^{t*}_{m_t} + \sum_{r=1}^{R} \beta^t_{\cdot r} \cdot y^{t*}_{m_t} \cdot z^{cov}_{ir}.$$

Here, $y^{t*}_{m_t}$ is the score assigned to category m_t of the tth indicator.

With categorical indicators and local dependencies, we get

$$\eta^h_{\mathbf{m}|x,\mathbf{z}_i} = \sum_{t\in h} \left\{ \beta^t_{m_t 0} + \beta^t_{m_t x0} + \sum_{r=1}^{R} \beta^t_{m_t r} \cdot z^{cov}_{ir} \right\} + \sum_{t\in h, t'\in h, t<t'} \beta^{tt'}_{m_t m_{t'}},$$

where the term $\beta^{tt'}_{m_t m_{t'}}$ captures the association between indicators t and t' in set h. For ordinal indicators, we impose the appropriate constraints using the category scores $y^{t*}_{m_t}$.

The linear term in the multinomial logit model for the latent classes equals

$$\eta_{x|\mathbf{z}_i} = \gamma_{x0} + \sum_{r=1}^{R} \gamma_{xr} \cdot z^{cov}_{ir}.$$

The intercept parameters γ_{x0} and the slope parameters γ_{xr} are subjected to the appropriate identifying constraints; that is, either $\sum_{x=1}^{K} \gamma_{xr} = 0$, $\gamma_{1r} = 0$, or $\gamma_{Kr} = 0$, for $0 \le r \le R$.

3.2 The Standard LC Model for Categorical Indicators

The standard LC model as described by Goodman (1974a, 1974b), Haberman (1979), and Lazarsfeld and Henry (1968) is a LC Cluster model containing only categorical indicators.[8] An example of a LC model with three categorical indicators ($T = 3$) is

$$P(y_{i1} = m_1, y_{i2} = m_2, y_{i3} = m_3) = \sum_{x=1}^{K} P(x) \prod_{t=1}^{3} P(y_{it} = m_t|x). \qquad (7)$$

where

$$\prod_{t=1}^{3} P(y_{it} = m_t|x) = P(y_{i1} = m_1|x)\, P(y_{i2} = m_2|x) P(y_{i3} = m_3|x)$$

As can be seen from this probability structure, the indicators y_{i1}, y_{i2}, and y_{i3} are assumed to be mutually independent given that one belongs to a certain latent class. This conditional independence constraint is sometimes referred to as the local independence assumption. Note that – using the terminology introduced above – each of the H sets consists of a single y variable, or equivalently $H = T$.

The conditional (response) probabilities $P(y_{it} = m|x)$ are parameterized as follows:

$$P(y_{it} = m|x) = \frac{\exp(\eta^t_{m|x})}{\sum_{m'=1}^{M_t} \exp(\eta^t_{m'|x})}\,,$$

with

$$\eta^t_{m|x} = \beta^t_{m0} + \beta^t_{mx0}.$$

If the corresponding indicator is a nominal variable, except for the identifying effect or dummy coding constraints, we do not need to impose further restrictions on the β^t_{mx0} parameters. On the other hand, if y_{it} is an ordinal indicator, the two-variable term appearing in the logistic form of $P(y_{it} = m|x)$ is restricted using the category scores y^{t*}_m; that is,

$$\beta^t_{mx0} = \beta^t_{x0} \cdot y^{t*}_m.$$

[8] Textbooks and introductory papers on the standard LC model include Bartholomew and Knott (1999), Clogg (1995), Dayton (1998), Dillon and Kumar (1994), Hagenaars (1990, 1993), Hagenaars and McCutcheon (2002), Heinen (1996), Magidson and Vermunt (2004), McCutcheon (1987), Shockey (1988), Vermunt (1997), and Vermunt and Magidson (2004).

Such a restricted association term is sometimes referred to as a row- or column-association model, depending on whether we see the latent variable as the row or the column variable (Agresti, 2002; Goodman, 1979). As was explained in the previous section, this yields an adjacent-category ordinal logit model for response variable y_{it}.

3.3 Covariates

An important extension of the standard LC model described above is obtained by the possibility of including covariates (Clogg, 1981; Dayton and McReady, 1988; Hagenaars, 1990, 1993).[9] Inclusion of covariates to predict class membership is straightforward within the general framework of the model defined in equation (1).

Suppose we have a model with three categorical indicators and two covariates (z_{i1}^{cov} and z_{i2}^{cov}). The LC Cluster model for this situation is

$$P(y_{i1}=m_1, y_{i2}=m_2, y_{i3}=m_3|z_{i1}^{cov}, z_{i2}^{cov}) = \sum_{x=1}^{K} P(x|z_{i1}^{cov}, z_{i2}^{cov}) \cdot \prod_{t=1}^{3} P(y_{it}=m_t|x). \quad (8)$$

Note that compared to the model without covariates described in equation (7), we replaced $P(x)$ by $P(x|z_{i1}^{cov}, z_{i2}^{cov})$, which makes the distribution of x dependent on z_{i1}^{cov} and z_{i2}^{cov}. It is important to be aware of the fact that we are making an additional set of conditional independence assumptions: The indicators are assumed to be independent of the covariates given the latent variable x.

The probability $P(x|z_{i1}^{cov}, z_{i2}^{cov})$ is restricted by means of a multinomial logistic regression model to exclude higher-order interaction terms, as well as to be able to deal with numeric (ordinal, discrete interval, or continuous) covariates. This yields

$$P(x|z_{i1}^{cov}, z_{i2}^{cov}) = \frac{\exp(\eta_{x|z_{i1},z_{i2}})}{\sum_{x'=1}^{K}\exp(\eta_{x'|z_{i1},z_{i2}})}, \quad (9)$$

with

$$\eta_{x|z_{i1},z_{i2}} = \gamma_{x0} + \gamma_{x1}\, z_{i1} + \gamma_{x2}\, z_{i2}.$$

[9] Other references on this topic are Dayton (1998), McCutcheon (1988, 1994), Van der Heijden, Dessens, and Böckenholt (1996), and Vermunt (1997).

As can be seen, the three-variable interaction is not included in the logit model. To simplify the model, Latent GOLD always excludes higher-order interactions between covariates from a model.[10] As was explained before, for nominal covariates, the program sets up the appropriate effects or dummies. This means that z_{i1}^{cov} and z_{i2}^{cov} could also be the two effects or dummies corresponding to a three-category nominal covariate.

We call this procedure for including covariates in a model the "active covariates method": Covariates are active in the sense that the LC Cluster solution with covariates can be somewhat different from the solution without covariates. An alternative method, labeled "inactive covariates method", involves computing descriptive measures for the association between covariates and the latent variable after estimating a model without covariates. More detail on the latter method is given in the subsection explaining the ProbMeans output.

Another alternative is to use a three-step approach; that is, to estimate the LC model without covariates (step 1), obtain the class assignments (step 2), and estimate the covariate effects using the assigned class memberships (step 3). A bias adjusted step-three analysis is implemented in the Latent GOLD Step3 submodule.

3.4 Local Dependencies

As mentioned above, the local independence assumption is the basic assumption of the standard LC model. Lack of fit of a LC model is caused by violation of this assumption. The usual way to proceed is to increase the number of classes until a model with an acceptable fit is obtained. An alternative model fitting strategy that we would like to propagate is to relax the local independence assumption by allowing for associations between indicators, as well as direct effects of covariates on the indicators (Hagenaars, 1988; Vermunt, 1997). Latent GOLD calculates bivariate z–y and y–y residuals which can be used to detect which pairs of observed variables are more strongly related than can be explained by the formulated model.

As in the previous subsection, we will use an example of a LC model with three indicators and two covariates. Suppose that we would like to relax two local independence assumptions by assuming that y_{i1} and y_{i2} are directly

[10] Note that higher-order interactions can be included in a model by adding the corresponding product terms to the data base and using these as additional covariates.

related and that y_{i3} is affected by z_{i2}. This would modify the right-hand side of equation (8) as follows:

$$\sum_{x=1}^{K} P(x|z_{i1}^{cov}, z_{i2}^{cov}) P(y_{i1} = m_1, y_{i2} = m_2|x) P(y_{i3} = m_3|x, z_{i2}^{cov}).$$

The dependent variables y_{i1} and y_{i2} now serve as a joint dependent variable and y_{i3} is allowed to depend on z_{i2}^{cov}. The logit model for $P(x|z_{i1}^{cov}, z_{i2}^{cov})$ is the same as above. The linear term in the logit model for $P(y_{i1} = m_1, y_{i2} = m_2|x)$ equals

$$\eta_{m_1 m_2|x}^{12} = \beta_{m_1 0}^{1} + \beta_{m_1 x 0}^{1} + \beta_{m_2 0}^{2} + \beta_{m_2 x 0}^{2} + \beta_{m_1 m_2}^{12},$$

and, the term for $P(y_{i3} = m_3|x, z_{i2}^{cov})$ equals

$$\eta_{m_3|x,z_{i2}}^{3} = \beta_{m_3 0}^{3} + \beta_{m_3 x 0}^{3} + \beta_{m_3 2}^{3} z_{i2}^{cov}.$$

These linear terms are used again to exclude higher-order interactions from the model, as well as to use the information on the scale type of the variables.

Latent GOLD starts by setting up a probability structure corresponding to a local independence model, such as those described in equations (7) and (8). When users include local dependencies using information on bivariate residuals, the program automatically sets up the correct and most parsimonious probability structure for the situation concerned.

3.5 Finite Mixture Models for Continuous Response Variables

The Cluster submodule can not only be used to specify cluster-type models for categorical indicators, but also to estimate models with continuous indicators (Vermunt and Magidson, 2002). The basic structure of a LC Cluster model for continuous y variables is:[11]

$$f(\mathbf{y}_i) = \sum_{x=1}^{K} P(x)\, f(\mathbf{y}_i|x),$$

where different variants can be obtained by means of the specification of $f(\mathbf{y}_i|x)$. The least restrictive model is obtained by assuming that the y's come from class-specific multivariate normal distributions

$$f(\mathbf{y}_i|x) = (2\pi)^{-K_m/2} \left|\boldsymbol{\Sigma}_x\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\mathbf{y}_i - \boldsymbol{\mu}_x\right)' \boldsymbol{\Sigma}_x^{-1} \left(\mathbf{y}_i - \boldsymbol{\mu}_x\right)\right\}$$

[11] Note that we dropped the index h because all variables belong to the same set ($H = 1$).

This model is also known as an unrestricted FM of multivariate normals (Banfield and Raftery, 1993; McLachlan and Basford, 1988; McLachlan and Peel, 2000; Wolfe, 1970). As can be seen, each latent class (or mixture component) has its own set of means $\boldsymbol{\mu}_x$ and its own variance-covariance matrix $\boldsymbol{\Sigma}_x$.

Specification of more restricted models for continuous y variables typically involves fixing some off-diagonal elements of the covariance matrix to zero. The most restrictive model assumes that all covariances equal zero, which is equivalent to the local independence assumption. This model can also be written as

$$f(\mathbf{y}_i) = \sum_{x=1}^{K} P(x) \prod_{t=1}^{T} f(y_{it}|x), \tag{10}$$

with

$$f(y_{it}|x) = \frac{1}{\sqrt{2\pi\sigma_{t,x}^2}} \exp\left\{ -\frac{1}{2} \frac{(y_{it} - \mu_{t,x})^2}{\sigma_{t,x}^2} \right\}.$$

Note that the probability structure in equation (10) is similar to the one of the standard LC model for categorical variables described in equation (7).

Intermediate models are be obtained by setting some but not all off-diagonal elements of the $\boldsymbol{\Sigma}_x$ matrices to zero, yielding specifications in which some pairs of indicators are mutually independent within classes whereas others are not. As was explained in Section 2, in such situations, we will work with sets of y variables, where indicators belonging to different sets are assumed to be locally independent and indicators belonging to the same set may be correlated. As with categorical y variables, the default is the local independence model described in equation (10). Users can include local dependencies one by one, and the program subsequently sets up the correct and most parsimonious probability structure.

An important issue in the specification of mixtures of normal distributions is whether to work with Cluster-dependent or Cluster-independent error variances and covariances. So far, we assumed that the error variances and covariances were Cluster dependent. However, when the number of y variables and/or number of latent classes is large, this may yield models that have many parameters to be estimated. By replacing $\sigma_{t,x}^2$ by σ_t^2 or $\boldsymbol{\Sigma}_x$ by $\boldsymbol{\Sigma}$ in the above formulas, one obtains more parsimonious structures with Cluster-independent variances and covariances. It is also possible to allow only the variances to differ across Clusters, but assume that the covariances are the same for all Clusters.

Finally, as in LC Cluster models for categorical variables, it is possible to include covariates as predictors of class membership in mixtures of normals. Covariates can not only influence class membership, but can also have direct effects on the Cluster-specific means. The most general finite mixture of multivariate normals with covariates is defined in equation (6), with Cluster-specific densities of the form described in equation (3).

3.6 LC Cluster Models for Mixed Mode Data

The most general LC Cluster model is the model for mixed mode data (Everitt, 1988; Hunt and Jorgensen, 1999; Lawrence and Krzanowski, 1996; Moustaki, 1996; Vermunt and Magidson, 2002). This model is used when one has y variables of different scale types. The structure that serves as the starting point is again the local independence structure that we also used for categorical and continuous variables (see equation 5). For each indicator, the user has to specify whether it is nominal, ordinal, continuous, or a count. As in the above models for categorical and continuous indicators, it is possible to include covariates in LC Cluster models for mixed mode data. These covariates can also have direct effects on the various types of indicators.

Local dependencies between pairs of categorical (nominal or ordinal) variables and between pairs of continuous variables are dealt with in the same way as discussed above; that is, via joint multinomial and multivariate normal distributions. Currently, there is no option for including other kinds of y–y association. There is, however, an indirect method (a trick) to specify a local dependency between, for instance, a categorical indicator, say y_{i1}, and a continuous indicator, say y_{i2} (Vermunt and Magidson, 2002). This can be accomplished by duplicating the categorical indicator and using it both as a covariate and as an indicator in the model of interest. The local dependency is obtained by specifying that the "covariate y_{i1}" has a direct effect on y_{i2}, but does not affect the latent variable x. This yields the conditional Gaussian distribution for y_{i2} proposed by Hunt and Jorgensen (1999). In this way, one can specify direct effects of continuous on categorical and count indicators, of categorical on continuous and count indicators, and of counts on categorical, continuous and other count indicators.

3.7 Parameter Restrictions in Cluster Models

Several types of restrictions can be imposed on the parameters of a LC Cluster model. These are:

- Equal x–y effects across indicators of the same scale type. In the example with three nominal indicators described above, assuming that the three indicators have same numbers of categories, this constraint would imply that $\beta^1_{mx0} = \beta^2_{mx0} = \beta^3_{mx0}$. This restriction is especially useful in scaling and IRT-like applications. With dichotomous indicators, for example, it yields a LC or nonparametric Rasch model and with ordinal indicators a LC or nonparametric partial-credit model (Heinen, 1996; Vermunt, 2001).

- Order-restricted Clusters. With this constraint, cluster-specific item probabilities and means will be restricted to be monotonically increasing. It yields what is usually referred to as ordinal LC analysis (Croon, 1990, 2002; Vermunt, 2001). For indicators which are specified to be ordinal, continuous, or counts, the order-restricted-clusters constraint implies the following inequality on the regression parameters: $\beta^t_{x0} \leq \beta^t_{x+1,0}$, for $1 \leq x \leq K-1$. In other words, the parameter corresponding to class $x+1$ should be at least as large as the parameter corresponding to class x. With nominal indicators the constraint is on the adjacent-category logits $\beta^t_{m+1,x0} - \beta^t_{mx0}$. More precisely, the inequality $\beta^t_{m_t+1,x0} - \beta^t_{m_t x0} \leq \beta^t_{m_t+1,x0} - \beta^t_{m_t x+1,0}$, for $1 \leq x \leq K-1$ and $1 \leq m_t \leq M_t - 1$, yields Croon's ordinal LC model, which is strongly connected to nonparametric IRT (see Vermunt, 2001).[12]

- Exclude certain x–y and z–x effects. The possibility to equate such an effect to zero can be used to test its significance using a likelihood-ratio test.

- Cluster independent variances/covariances. Variances of and covariances between continuous response variables can be restricted to be equal across Clusters, yielding a simpler model-based clustering model.

[12] In Croon's original formulation, the inequality restrictions were imposed on cumulative logits rather than on adjacent-category logits (see Vermunt 2001 for the various variants). In the dichotomous case, the two are, of course, equivalent.

- The cluster-specific error variances of the continuous response variables are restricted to be at least 1.0e-6 times the observed variances in the sample.

4 DFactor Models

The LC DFactor model implemented in Latent GOLD is a model with:

1. L dichotomous/ordinal latent variables called discrete factors (DFactors), which may be assumed to be mutually dependent or independent,
2. T response variables y_{it} (indicators) that can be nominal, ordinal, continuous, and/or counts,
3. certain DFactor-indicator effects ("loadings") restricted to zero,
4. R numeric and nominal covariates z_{ir}^{cov} affecting the x_ℓ,
5. direct relationships between indicators and/or direct effects of covariates on indicators.

The main difference between a DFactor and a LC Cluster model is that the former may contain more than one latent variable. Another difference is that in the DFactor model the categories of the latent variables are assumed to be ordered. Thus, rather than working with a single nominal latent variable, here we work with one or more dichotomous or ordered polytomous latent variables (Magidson and Vermunt, 2001; Vermunt and Magidson, 2005a). The advantage of this approach is that it guarantees that each of the DFactors is one-dimensional.

The primary difference between our LC DFactor model and the traditional factor-analytic model is that the latent variables (DFactors) are assumed to be dichotomous or ordinal as opposed to continuous and normally distributed. Because of the strong similarity with traditional factor analysis, we call this approach LC DFactor analysis. There is also a strong connection between DFactor models and IRT or latent trait models. Actually, DFactor models are discretized variants of well-known latent trait models for dichotomous and polytomous items (Heinen, 1996; Vermunt, 2001; Vermunt and Magidson, 2005a). [13]

[13] Two edited volumes paying special attention to the relationship between latent trait and LC models are Langeheine and Rost (1988) and Rost and Langeheine (1997).

As in maximum likelihood factor analysis, modeling under the LC DFactor approach can proceed by increasing the number of discrete factors until a good fitting model is achieved. This approach to LC modeling provides a general alternative to the traditional method of obtaining a good fitting model by increasing the number of latent classes. In particular, when working with dichotomous uncorrelated DFactors, there is an exact equivalence in the number of parameters of the two models. A DFactor model with 1 discrete factor has the same number of parameters as a 2-class LC Cluster model, a model with 2 discrete factors as a 3-class model, a model with 3 discrete factors as a 4-class model, etc. Thus, in an exploratory analysis, rather than increasing the number of classes one may instead increase the number of DFactors until an acceptable fit is obtained.

4.1 Probability Structure and Linear Predictors

The most general probability structure one can use in a DFactor analysis with L discrete factors is

$$\begin{aligned} f(y_i|z_i^{cov}) &= \sum_{x_1=1}^{K_1}\sum_{x_2=1}^{K_2}\cdots\sum_{x_L=1}^{K_L} P(x_1,x_2,\ldots,x_L|z_i^{cov}) \\ &\quad \cdot \prod_{h=1}^{H} f(y_{ih}|x_1,x_2,\ldots,x_L,z_i^{cov}). \end{aligned}$$

The similarity of this equation with the LC Cluster model described in equation (6) becomes clearer by noting that the joint latent variable $x_1, x_2, \ldots, x_L$ can also be considered to be a single latent variable with $K = \prod_{\ell=1}^{L} K_\ell$ categories. A 3-DFactor model with 2 levels per DFactor, for example, is in fact a restricted 8-Cluster model. Note that as Cluster models, DFactor models may contain direct effects between indicators and direct effects of covariates on indicators.

Because of the similarity between the linear predictors in LC DFactor models to those in Cluster models, it is not necessary to provide again the detailed formulas of the various linear predictors. Instead we concentrate on the three main differences:

1. In an L-DFactor model there is not one but L terms describing the relationship between the latent variables and the indicators/covariates.

2. DFactors are not nominal but ordinal variables, and are handled in a manner similar to ordinal indicators: Equal distance scores between 0 and 1 are assigned to the K_ℓ categories of DFactor ℓ. With $K_\ell = 3$, for example, the scores would be $x_1^{\ell*} = 0$, $x_2^{\ell*} = 0.5$, and $x_3^{\ell*} = 1$, for categories 1, 2, and 3, respectively. As is shown below, these scores are used to restrict the effect parameters involving a DFactor.

3. Because there is more that one latent variable, the model specification also involves the associations between these discrete latent factors.

To illustrate the impact of the first two differences, let us have a look at the linear term for continuous indicators and counts, which equals

$$\eta_{x,\mathbf{z}_i}^t = \beta_0^t + \sum_{\ell=1}^{L} \beta_{\ell 0}^t \cdot x_{x_\ell}^{\ell*} + \sum_{r=1}^{R} \beta_r^t \cdot z_{ir}^{cov},$$

where $\beta_{\ell 0}^t$ is the effect of DFactor x_ℓ on indicator y_{it}. It can easily be seen that there is a separate $\beta_{\ell 0}^t$ term for each DFactor and that the numeric category scores $x_{x_\ell}^{\ell*}$ appear in the regression equation. Whereas for ordinal indicators the term $\beta_{\ell 0}^t$ has the same form as for continuous indicators and counts, for nominal indicators it becomes $\beta_{m_t \ell 0}^t$, where the extra index m_t refers to the category of y_{it}.

The linear term in the regression model for the DFactors equals

$$\eta_{x|\mathbf{z}_i} = \sum_{\ell=1}^{L} \gamma_{x_\ell 0}^{\ell} + \sum_{\ell=1}^{L} \sum_{r=1}^{R} \gamma_{\cdot r}^{\ell} \cdot x_{x_\ell}^{\ell*} \cdot z_{ir}^{cov} + \sum_{\ell=1}^{L} \sum_{\ell'=1}^{\ell-1} \gamma_{\cdot\cdot}^{\ell\ell'} \cdot x_{x_\ell}^{\ell*} \cdot x_{x_{\ell'}}^{\ell'*},$$

where $\gamma_{x_\ell 0}^{\ell}$ are the intercept terms and $\gamma_{\cdot r}^{\ell}$ the covariate effects for DFactor ℓ, and $\gamma^{\ell\ell'}$ is the parameter capturing the association between discrete factors ℓ and ℓ'. Note that this is a multivariate variant of the adjacent-category ordinal logit model.

4.2 A Two-DFactor Model for Nominal Indicators

To illustrate the DFactor model, let us assume that we have a two-DFactor model for three nominal indicators. The corresponding probability structure is of the form

$$P(y_{i1} = m_1, y_{i2} = m_2, y_{i3} = m_3) = \sum_{x=1}^{K} P(x_1, x_2) \prod_{t=1}^{3} P(y_{it} = m_t | x_1, x_2).$$

The conditional response probabilities $P(y_{it} = m_t|x_1, x_2)$ are restricted by means of logit models with linear terms

$$\eta^t_{m|x_1,x_2} = \beta^t_{m0} + \beta^t_{m10} \cdot x^{1*}_{x_1} + \beta^t_{m20} \cdot x^{2*}_{x_2}.$$

As can be seen, two-variable terms are restricted using the category scores $x^{\ell*}_{x_\ell}$ and higher-order interaction terms are excluded from the model. As a results of these two types of constraints, the parameters describing the strength of relationships between the DFactors and the indicators – here, β^t_{m10} and β^t_{m20} – can be interpreted as factor loadings.

In the standard DFactor model, the discrete factors are specified to be dichotomous, which means that the row-association structures are not real constraints. An important extension of this standard model is that the number of levels of a DFactor can be increased, which makes it possible to describe the distribution of the underlying factor more precisely. Note that the levels of the DFactors remain ordered by the use of fixed equal-interval category scores in their relationships with the indicators.[14] Because of this parameterization, each additional level costs only one degree of freedom; that is, there is one additional class size to be estimated.

In the default setting, the discrete factors are assumed to independent of one another. This is specified by the appropriate logit constraints on the latent probabilities. In the two-DFactor case, this involves restricting the linear term in the logit model for $P(x_1, x_2)$ by

$$\eta_{x_1,x_2} = \gamma^1_{x_10} + \gamma^2_{x_20}.$$

Working with correlated DFactors is comparable to performing an oblique rotation. The association between each pair of discrete factors is described by a single uniform association parameter:

$$\eta_{x_1,x_2} = \gamma^1_{x_10} + \gamma^2_{x_20} + \gamma^{12}_{\cdot\cdot} \cdot x^{1*}_{x_1} \cdot x^{2*}_{x_2}.$$

It should be noted that contrary to traditional factor analysis, the LC DFactor model is identified without additional constraints, such as setting certain DFactor loadings equal to zero.[15] Nevertheless, it is possible to specify models in which DFactor loadings are fixed to zero. Together with the possibility

[14] Several authors have described such a scoring of latent classes. See, for instance, Clogg (1988), Formann (1992), Formann and Kolhmann (1998), Heinen (1996), Uebersax (1993, 1997), and Vermunt (2001).

[15] Of course, this is only true if there are sufficient indicators compared to the number of DFactors.

to include DFactor correlation in the model, this option can be used for a confirmatory factor analysis.

4.3 Other Possibilities

Above we presented a two-DFactor model for nominal categorical variables. We discussed several important extensions of the standard model, such as increasing the number of categories of the latent variables, assuming discrete factors to be correlated, and setting DFactor loadings equal to zero.

Other extended possibilities are similar to what we discussed in the context of the LC Cluster model. Indicators can not only be nominal but also ordinal, continuous, or counts. In addition, as in the Cluster models described in the previous section, one can specify models with local dependencies, as well as models with nominal and numeric covariates.

4.4 Parameter Restrictions in DFactor Models

Several types of restrictions can be imposed on the parameters of a DFactor model. These are:

- Equal x_ℓ–y effects for DFactor ℓ across indicators of the same type. In the example of a two-DFactor model with three nominal indicators, this constraint would, for example, imply that $\beta^1_{m10} = \beta^2_{m10} = \beta^3_{m10}$ for the first DFactor. This restriction is especially useful in scaling and IRT-like applications. With dichotomous indicators it yields a semi-parametric Rasch model and with ordinal indicators a semi-parametric partial-credit model (Heinen, 1996; Vermunt, 2001).

- Excluding certain x_ℓ–y and z–x_ℓ effects from the model. The option to exclude certain x_ℓ–y terms makes it possible to indicate that discrete factor ℓ is related to some but not all indicators, which is required in confirmatory models. Excluding a term can also be used to test its significance using a likelihood-ratio test.

- Include/Exclude DFactor-DFactor associations (default=exclude). Associations between discrete factors can either be included or excluded. As in standard factor analysis, one will typically exclude these DFactor-DFactor associations in an exploratory analysis, but include them in a confimatory analysis.

- The error variances of continuous indicators are restricted to equal at least 1.0e-6 times the observed variances in the sample.

5 Latent Class Regression Models

The LC or FM Regression model implemented in Latent GOLD is a model with:[16]

1. a single nominal latent variable x,
2. T_i replications or repeated observations of a single dependent variable y_{it}, which may be nominal, ordinal, continuous, or a binomial or Poisson count,
3. Q numeric or nominal predictors z_{itq}^{pred} affecting y_{it} via a GLM, where parameters may differ across latent classes,
4. zero, equality, fixed value, and order restrictions on regression coefficients,
5. R numeric or nominal covariates z_{ir}^{cov} affecting x.

The main difference between LC Regression analysis and the other forms of LC analysis implemented in Latent GOLD is that it contains a single dependent variable, which may, however, be observed more than once for each case. These multiple responses may be experimental replications, repeated measurements at different time points or occasions, clustered observations, responses on a set of questionnaire items, or other types of dependent observations. The value of the dependent variable for case i at replication t is denoted by y_{it}, and its total number of replications by T_i. Note that the index i in T_i makes it possible to deal with unequal numbers of observations per case.

In the context of LC Regression analysis, it makes sense to make a distinction between two types of exogenous variables:

1. variables influencing the latent variable, which we call covariates,

[16] References to LC or FM Regression analysis are Agresti (2002, section 13.2), Vermunt and Van Dijk (2001), Wedel and DeSarbo (1994), and Wedel and Kamakura (1998).

2. variables influencing the dependent variable, which we call predictors.

Covariates will be denoted by z_{ir}^{cov} and predictors by z_{itq}^{pred}, where the index t in z_{itq}^{pred} reflects that the value of a predictor may change across replications. A covariate, on the other hand, has the same value across all replications of a particular case.

Note that, in fact, we are dealing with a two-level data set, where t indexes the lower-level observations within higher-level observation i. Covariates serve as higher-level exogenous variables and predictors as lower-level exogenous variables. This illustrates that Latent GOLD can be used to estimate (nonparametric) two-level or random-coefficient models (Aitkin, 1999; Skrondal and Rabe-Hesketh, 2004). Using t as an index for time points or time intervals, one obtains nonparametric random-coefficients models for longitudinal data, such growth data, event history data, and panel models with (Vermunt, 1997, 2002a, 2007; Vermunt and Van Dijk, 2002; Wedel et al., 1995; Zackin, De Gruttola, and Laird, 1996).

5.1 Probability Structure and Linear Predictors

The most general probability structure that can be used in the Regression submodule takes on the following form:

$$f(\mathbf{y}_i|\mathbf{z}_i^{cov},\mathbf{z}_i^{pred}) = \sum_{x=1}^{K} P(x|\mathbf{z}_i^{cov}) \prod_{t=1}^{T_i} f(y_{it}|x,\mathbf{z}_{it}^{pred}) \,. \tag{11}$$

The main differences from the Cluster submodule (see equation 6) are that in Regression we make a distinction between covariates and predictors, we allow for different numbers of replications per case, we assume that the conditional densities $f(y_{it}|x,\mathbf{z}_{it}^{pred})$ have the same form for each t, and we do not allow for direct effects between the multiple responses.

In Latent GOLD's Regression submodule, it is possible to specify a *replication weight* v_{it} for each of the records of case i. This modifies the definition of $f(\mathbf{y}_i|\mathbf{z}_i^{cov},\mathbf{z}_i^{pred})$ somewhat:

$$f(\mathbf{y}_i|\mathbf{z}_i^{cov},\mathbf{z}_i^{pred}) = \sum_{x=1}^{K} P(x|\mathbf{z}_i^{cov}) \prod_{t=1}^{T_i} \left\{ f(y_{it}|x,\mathbf{z}_{it}^{pred}) \right\}^{v_{it}} . \tag{12}$$

Two possible applications of replication weights include grouping of records and differential weighting of responses. Another possible application is the

analysis of multinomial counts. With an M-category multinomial count response variable, one would get M records for each case with responses equal to $1, 2, 3, ..., M$, respectively, and replication weights equal to the number of times the category concerned was selected.

Another slight modification of the probability structure defined in equation (11) occurs with the *zero-inflated* option. Depending on the scale type, the model is expanded with either one or M latent classes that are assumed to give a particular response with probability one. Zero-inflated Poisson and binomial count models are obtained by adding a single latent class to the model, yielding a model with $K + 1$ Classes. Class $K + 1$ has a zero Poisson rate or zero binomial probability, or, equivalently, a probability of one of having zero events; that is, $P(y_{it} = 0|x, \mathbf{z}_{it}^{pred}) = 1$ for $x = K + 1$. The same happens in the case of a continuous dependent variable with the additional modification that the dependent variable is assumed to be censored-normal distributed in the other K Classes. This model is sometimes referred to a censored-inflated linear regression model. For ordinal and nominal dependent variables, M Classes are added to the model, each of which responds with probability one to a certain category; that is, $P(y_{it} = m|x, \mathbf{z}_{it}^{pred}) = 1$ for $x = K + m$. Such Classes are sometimes referred to a stayer Classes (in mover-stayer models) or brand-loyal Classes (in brand-loyalty models).

Because the linear predictor in $P(x|\mathbf{z}_i^{cov})$ has the same form as in LC Cluster models (see equation 9), we will pay no further attention to it here. The other two linear predictors of interest – $\eta_{x,\mathbf{z}_{it}}$ and $\eta_{m|x,\mathbf{z}_{it}}$ – require more explanation because these have a specific form in LC Regression models. For *continuous and count responses*, we use

$$\eta_{x,\mathbf{z}_{it}} = \beta_{x0} + \sum_{q=1}^{Q} \beta_{xq} \cdot z_{itq}^{pred},$$

where β_{x0} is a Class-specific intercept and β_{xq} the Class-specific regression coefficient corresponding to predictor number q. Depending on the scale type of the response variable, this yields either a standard linear, log-linear Poisson, or binary logistic regression model with parameters that differ across latent classes. The above linear term differs from the one used in LC Cluster models in that all parameters can vary between Classes and no further restrictions are needed for identification. Another important difference is that neither η or β has an index t, which implies that parameters do not differ across repeated responses. Differences in predicted values between the

multiple responses can thus only be the result of varying predictor values.[17]

For *nominal responses*, we use a multinomial logit model having a linear predictor of the form

$$\eta_{m|x,\mathbf{z}_{it}} = \beta_{xm0} + \sum_{q=1}^{Q} \beta_{xmq} \cdot z_{itq}^{pred},$$

with – depending on the selected coding type for nominal variables – the identification restrictions $\sum_{m=1}^{M} \beta_{xmq} = 0$, $\beta_{x1q} = 0$, or $\beta_{xMq} = 0$, for $0 \leq q \leq Q$. An *ordinal response variable* is modeled with an adjacent-category logit model in which

$$\eta_{m|x,\mathbf{z}_{it}} = \beta_{xm0} + \sum_{q=1}^{Q} \beta_{x\cdot q} \cdot y_m^* \cdot z_{itq}^{pred}.$$

Here, y_m^* is the score assigned to category m of the response variable. In the ordinal case, we need only an identifying effect or dummy constraints for the intercept term β_{xm0}.

5.2 Some Special Cases

The simplest probability structure for a LC "Regression" model occurs if there is a single response per case and no predictors; that is,

$$f(y_i) = \sum_{x=1}^{K} P(x) f(y_i|x)\,.$$

This yields a simple univariate finite mixture model in which the mean and possibly also the variance of the distribution of y_i is assumed to be Class-dependent. Such a model without predictors makes it possible to describe the unobserved heterogeneity with respect to the distribution of y_i in the population under study.[18]

A more useful LC Regression model is obtained by including predictors in the model, such as,

$$f(y_i|z_{i1}^{pred}, z_{i2}^{pred}) = \sum_{x=1}^{K} P(x) f(y_i|x, z_{i1}^{pred}, z_{i2}^{pred})\,.$$

[17] By defining an independent variable "replication number" or "time", as well as the necessary product terms, one can make parameters replication or time dependent.

[18] Relevant references on this topic are Böhning (2000), Dillon and Kumar (1994), Everitt and Hand (1981), Laird (1978), McLachlan and Basford (1988), McLachlan and Peel (2000), Magidson and Vermunt (2003a), and Titterington, Smith, and Makov (1985).

Here, $f(y_i|x, z_{i1}^{pred}, z_{i2}^{pred})$ denotes the distribution of the dependent variable y given a person's class membership x and predictor values z_{i1}^{pred} and z_{i2}^{pred}. Depending on the type of dependent variable, the expected value of the appropriate distribution is restricted by means of a logistic, log-linear, or linear regression model.[19]

An important extension of the above LC Regression model is obtained by making class membership dependent on covariates (Kamakura, Wedel, and Agrawal, 1994; Vermunt, 1997). An example of such a model is:

$$f(y_i|z_{i1}^{covc} z_{i2}^{cov}, z_{i1}^{pred}, z_{i2}^{pred}) = \sum_{x=1}^{K} P(x|z_{i1}^{cov}, z_{i2}^{cov}) f(y_i|x, z_{i1}^{pred}, z_{i2}^{pred})$$

In this model, it is assumed that the probability of belonging to latent class x depends on the values of z_{i1}^{cov} and z_{i2}^{cov}. This is equivalent to the way covariates can be used in LC Cluster models.

As already mentioned above, there may be more than one observation per case; that is, there may be more than one replication of the same dependent and independent variables for each observational unit. Extending the above model to multiple replications yields the following probability structure:

$$f(\mathbf{y}_i|z_{i1}^{cov}, z_{i2}^{cov}, \mathbf{z}_{i1}^{pred}, \mathbf{z}_{i2}^{pred}) = \sum_{x=1}^{K} P(x|z_{i1}^{cov}, z_{i2}^{cov}) \prod_{t=1}^{T_i} f(y_{it}|x, z_{it1}^{pred}, z_{it2}^{pred}) \,.$$

Such a LC Regression model for repeated measures is very similar to multilevel (two-level), mixed, or random-coefficients models, in which random effects are included to deal with the dependent observations problem. The LC Regression model is, in fact, a nonparametric random-effects model (Agresti, 2002, section 13.2; Aitkin, 1999; Skrondal and Rabe-Hesketh, 2004; Vermunt and Van Dijk, 2001).

5.3 Restrictions for the Class-Specific Regression Coefficients

Various types of restrictions can be imposed on the Class-specific regression coefficients: intercepts and predictor effects can be fixed to zero, restricted to

[19] For some applications, see Böckenholt (1993), Land, McCall and Nagin (1996), Mare (1994), and Wedel and DeSarbo (1994).

be equal across certain or all Classes, and constrained to be ordered. Moreover, the effects of numeric predictors can be fixed to one. These constraints can either be used as a priori restrictions derived from theory or as post hoc restrictions on estimated models.

Certain restrictions apply to parameters within each Class, while others apply across Classes. The within-Class restrictions are:

- No Effect: the specified effect(s) are set to zero selected latent classes;
- Offset: the selected effect(s) are set to one, thus serving as an offset.[20] The offset effect applies to numeric predictors only, and only with ordinal, continuous, Poisson count, and binomial count responses (thus not with a nominal dependent variable).

Between-Class restrictions are:

- Merge Effects: the intercept or the effect a selected predictor is equated across 2 or more specified Classes;
- Class Independent: the intercept or the effects of a selected predictor is equated across all Classes;
- Order Restriction (ascending or descending): in each Class, the effect of a selected numeric predictor is assumed to have the same sign or the effects corresponding to a selected nominal predictor are assumed to be ordered (either ascending or descending). That is, for numeric predictors, the ascending restriction implies that the Class-specific coefficients should be at least zero ($\beta \geq 0$) and the descending restriction that they are at most zero ($\beta \leq 0$). For nominal predictors, ascending implies that the coefficient of category $p+1$ is larger than or equal to the one of category p ($\beta_p \leq \beta_{p+1}$, for each p) and descending that the coefficient of category $p+1$ is smaller than or equal to the one of category p ($\beta_p \geq \beta_{p+1}$, for each p). Similarly to what was explained in the context of the order-restricted Cluster model, with a nominal dependent

[20] The term offset stems from the generalized linear modeling framework. It refers to a regression coefficient that is fixed to 1, or equivalently, to a component that offsets the linear part of the regression model by a fixed amount. An offset provides the same role as a cell weight in log-linear an logit analysis. In log-linear and logit model, an offset is, in fact, the log of a cell weight.

variable, the imposed order restriction applies to each of the adjacent category logits. In the case of ascending, this implies $\beta_{m+1} - \beta_m \geq 0$ for numeric predictors and $\beta_{m+1,p} - \beta_p \leq \beta_{m+1,p+1} - \beta_{m,p+1}$ for nominal predictors (see Galindo and Vermunt, 2005; Vermunt, 1999; Vermunt and Hagenaars, 2004).

- The Class-specific error variances of a continuous dependent variable are restricted to be at least 1.0e-6 times the observed variance in the sample.

The "Class Independent" option can be used to specify models in which some coefficients differ across Classes while others do not. This can either be on a priori grounds or can be based on the test statistics from previously estimated models. More specifically, if the Wald(=) test is not significant, it makes sense to check whether an effect can be assumed to be Class independent.

There is a special variant of the Class-independent option called "No Simple" that can be used in conjunction with the intercept in an ordinal regression model. With this option, the intercept is modeled as $\beta_{xm0} = \beta_{\cdot m0} + \beta_{x\cdot 0} \cdot y^*_m$, where $\beta_{x\cdot 0}$ is subjected to an effect or dummy coding constraint. This specification of a Class-specific intercept is much more parsimonious and is, in fact, equivalent to how x-y relationships with ordinal y's are modeled in LC Cluster models. Rather that estimating $K \cdot M$ intercept terms, one now estimates only $M + K - 1$ coefficients; that is, one extra coefficient per extra latent class.

"Order Restrictions" are important if one has a priori knowledge about the sign of an effect. For example, the effect of price on persons' preferences is usually assumed to be negative – or better, non-positive – for each latent class (segment). If the price effect is specified to be "Descending", the resulting parameter estimate(s) will be constrained to be in agreement with this assumption. Another application of order restrictions occurs in growth modeling. Rather than assuming a specific functional form for the time dependence, one may wish to make the much less restrictive assumption that the time effect is monotonic (see, for example, Vermunt and Hagenaars, 2004).

The "No Effect" option makes it possible to specify a different regression equation for each latent class. More specifically, each latent class may have different sets of predictors affecting the responses. Post hoc constraints can be based on the reported z value for each of the coefficients. An example of

an a priori use of this constraint is the inclusion of a random-responder class, a latent class in which all coefficient are zero, except for the intercept. This is specified as follows:

	Class 1	Class 2	Class 3	Class 4	
Intercept	1	2	3	4	
Predictor1	–	2	3	4	,
Predictor2	–	2	3	4	

where "–" indicates that the effect is equal to 0. In this example, Class 1 is the random-responder class.

"Merge Effects" is a much more flexible variant of "Class Independent". It can be use to equate the parameters for any set of latent classes. Besides post hoc constraints, very sophisticated a priori constraints can be imposed with this option. An important application is the specification of DFactor-like structures in which each latent class corresponds to the categories of two or more latent variables. For example, consider a set of constraints of the form:

	Class 1	Class 2	Class 3	Class 4	
Intercept	1	1	3	3	
Predictor1	1	2	1	2	,
Predictor2	1	2	1	2	

where the numbers in the cells indicate "equal to Class #". This restricted 4-Class model is in fact a 2-dimensional DFactor model: the categories of DFactor 1 differ with respect to the intercept and the categories of DFactor 2 with respect to the two predictor effects. Specifically, level 1 of DFactor 1 is formed by Classes 1 and 2 and level 2 by Classes 3 and 4; level 1 of DFactor 2 is formed by Classes 1 and 3 and level 2 by Classes 2 and 4.

The option "Offset" can be used to specify any nonzero fixed-value constraint on the Class-specific effect of a numeric predictor in models for ordinal, continuous and count responses. This means that it is possible to refine the definition of any Class (segment) by enhancing or reducing the estimated numeric predictor effect for that Class. Recall that numeric predictor q enters as $\beta_{xq} \cdot z_{itq}^{pred}$ in the linear term of the regression model. Suppose, that after estimating the model, the estimate for β_{xq} turned out to be 1.5 for Class 1. If z_{itq}^{pred} is specified to be an offset, the effect of this predictor would be reduced (1.5 would be reduced to 1) for this Class. But suppose that you wish to enhance the importance of this predictor for Class 1; say, you wish to restrict

β_{xq} to be equal to 2. The trick is to recode the predictor, replacing each code by twice the value. Thus, the recoded predictor is defined as $2 \cdot z_{itq}^{pred}$. If we restrict the effect of this recoded predictor to 1, we obtain $1 \cdot 2 \cdot z_{itq}^{pred}$, which shows that the effect of z_{itq}^{pred} is equated to 2. Such recoding can be done easily within Latent GOLD, using the Replace option.

In addition to post hoc refinements to customize the definition of the resulting Classes, the offset restriction can also be used to make the latent classes conform to various theoretical structures. Probably the most important a priori application of "Offset" is the possibility of defining zero-inflated models for counts, which can also be specified with the special zero-inflated option discussed above. These are models containing one latent class that has a Poisson rate (binomial probability) equal to 0 and that is not affected by the other predictors. An example of a restrictions table corresponding to such a structure is:

	Class 1	Class 2	Class 3
Intercept	–	2	3
LargeNegative(-100)	*	–	–
Predictor1	–	2	3
Predictor2	–	2	3

.

Here, "–" means no effect and "*" means offset. As can be seen, Class 1 is only affected by an offset, and Classes 2 and 3 have there own intercept and predictor effects. The numeric "predictor" LargeNegative(-100) takes on the value -100 for all records.[21] As a result of the fixed effect of -100, the rate/probability of experiencing an event will be equal to zero. In the case of a Poisson count, the rate for someone belonging to Class 1 equals $\exp(-100) = 0$. In the binomial case, the probability that someone belonging to Class 1 experiences an event is $\exp(-100)/[1 + \exp(-100)] = 0$.

Now, we will discusses several more advanced applications of the restriction options. Suppose you assume that the effect of price is linear and negative (descending) for Classes 1-3 and unrestricted for Class 4. This can be accomplished by having two copies of the price variable in the model, say

[21] It is not necessary to assume that the -100 appears in all records. The value could also be -100 if a particular condition is fulfilled – for example, if the price of the evaluated product is larger than a certain amount – and 0 otherwise. This shows that the offset option provides a much more flexible way of specifying classes with zero response probabilities than the zero-inflated option.

Price1 and Price2. The effect of Price1 (numeric) is specified as ordered and is fixed to zero in Class 4. The effect of Price2 is fixed to zero in Classes 1-3. Similar tricks can be applied in LC growth models, in which one may wish to allow for a different type of time effect in each of the latent classes: for example, quadratic (time and time squared, both numeric) in Class 1, linear (time numeric) ascending in Class 2, unrestricted (time nominal) in Class 3, and no time effect in Class 4.

Suppose your assumption is that the effect of a particular predictor is at least 2. This can be accomplished by combining a fixed value constraint with an order constraint. More precisely, an additional predictor defined as $2 \cdot z_{itq}^{pred}$ is specified to be an offset and the effect of the original predictor z_{itq}^{pred} defined to be ascending.

Our final example is an exploratory variant of the DFactor structure described above. Suppose you want a two-DFactor model without assumptions on which discrete factor influences which predictor effects. This can be accomplished having 3 copies of all predictors plus and two extra intercepts (column with ones) in the data file. With two predictors (time and price), the restriction table is of the following form. The first copy (Intercept, Time1, and Price1) defines a main effect for each predictor. The second copy (Intercept2, Time2 and Price2) is used to define the first DFactor, a contrast between Classes 3/4 and 1/2. The third copy (Intercept3, Time3, and Price3) specifies DFactor 2 by means of a contrast between Classes 2/4 and 1/3.

6 Step-Three Analysis and Scoring

Latent GOLD 5.1 Basic contains a fourth submodule called Step3, which can be used for performing a bias-adjusted step-three analysis and for obtaining a scoring equation. We will first discuss three-step LC analysis and subsequently the scoring implementation.

6.1 Bias-Adjusted Step-Three Analysis

Three-step LC analysis means that the analysis of interest is performed using the following three steps:

1. A LC model is built for a set of response variables. This involves decisions regarding the indicators to be used, the number of classes needed, and other model features.

2. Using the final model from step 1, subjects are assigned to latent classes based on their posterior class membership probabilities $P(x|\mathbf{y}_i)$, and the class assignments are appended to the data file. Class assignment can be modal (to the class for which the posterior membership probability is largest) or proportional (to each class with a weight equal to the posterior membership probability for that class).

3. Using the assigned class memberships from step 2, the association between the class membership and external variables is investigated with simple cross-tabulations or ANOVAs, or with a multinomial logistic regression analysis. The external variables may be either predictors of class membership or (distal) outcomes influenced by class membership.

Bolck, Croon, and Hagenaars (2004) demonstrated that such a three-step approach underestimates the associations between external variables and class membership. They proposed resolving this problem by means of a specific correction method which involves modifying the third step. Latent GOLD 5.1 implements a generalized version of the adjustment method proposed by Bolck, Croon, and Hagenaars (2004), which we refer to as the BCH method (see Vermunt, 2010), as well as a second, conceptually simpler, adjustment method proposed by Vermunt (2010), which we refer to as the ML (maximum likelihood) method (see also Gudicha and Vermunt, 2013, and Bakk, Tekle, and Vermunt, 2013).

Let a_i denote the assigned class membership from step 2, x the true class membership from step 1, and $\mathbf{z}_i$ a set of covariates that one would like to use to predict x (Step3-Covariate option). Bolck et al. (2004) derived how the probability of interest, $P(x|\mathbf{z}_i)$, is related to the probability that we can model using the class assignments, $P(a_i|\mathbf{z}_i)$; that is,

$$P(a_i|\mathbf{z}_i) = \sum_{x=1}^{K} P(x|\mathbf{z}_i)P(a_i|x). \tag{13}$$

Note that this equation is in fact a LC model (with covariates) in which a_i serves as the single indicator of x, with conditional "response" probabilities $P(a_i|x)$. The latter represent the probability of being assigned to class a_i given that one belongs to class x. Note also that the $P(a_i|x)$ are, in fact, the entries of the Classification Table reported as part of the Classification Statistics, but rescaled to sum to 1 within rows (transformed to row proportions). The exact form of $P(a_i|x)$ depends on whether modal or proportional class assignment is used.

A similar equation can be set up for the situation in which the external variable is an outcome variable o_i (Step3-Dependent option); that is,

$$P(a_i, o_i) = \sum_{x=1}^{K} P(x)P(o_i|x)P(a_i|x). \tag{14}$$

This is a LC model with two indicators, the outcome variable o_i and the class assignments a_i. It should be noted that when selecting multiple dependent variables, a separate LC model will be set up or each dependent variable. The reason for this is that a Step3-Dependent analysis estimates the bivariate association between the classes and the outcome variable concerned.

The ML adjustment method involves estimating the LC model concerned by standard maximum likelihood. That is, it estimates the LC model in which $P(x|\mathbf{z}_i)$ is estimated freely with $P(a_i|x)$ fixed, or the LC model in which $P(o_i|x)$ is estimated freely with both $P(x)$ and $P(a_i|x)$ fixed. The quantities $P(x)$ and $P(a_i|x)$ are computed using the posterior class membership probabilities from step 2. When using proportional assignment, the ML adjusted step-three analysis requires expanding the data set to contain K records per individual with weights equal to the posterior membership probabilities. Because of this weighting, robust standard errors should be used (Bakk, Oberski, Vermunt, 2014).

In the BCH adjustment, instead of estimating a LC model, one performs the logistic regression analysis or the computation of the cross-tabulations or ANOVAs in the usual manner, with the difference that an expanded data file with K records per individual and a specific set of weights should be used. These weights are based on the inverse of the matrix with elements $P(a_i|x)$. Intuitively, what happens is that the observed information on a_i is transformed (reweighted) in such way so that it becomes information on the true class membership x. For more details, we refer the reader to Vermunt (2010), Gudicha and Vermunt (2013), and Bakk, Tekle, and Vermunt (2013).

Our simulation studies showed that the ML adjustment is the preferred approach when the external variables are covariates or categorical dependent variables (Vermunt, 2010; Bakk, Tekle, and Vermunt, 2013). However, with continuous dependent variables and counts, the BCH is the preferred method, which results from the fact that using the ML approach involves making strong additional assumptions about the distributions of the response variables (Bakk and Vermunt, 2016).

Lanza, Tan, and Bray (2013) recently proposed using distal outcomes as covariates rather than as dependent variables in the LC model. The Step3

submodule of Latent GOLD Basic can be used for a step-three variant of this approach; that is, for estimating a Step3-Covariate model in which the distal outcomes are used as covariates. The class-specific distal outcome probabilities/means will be reported in the Profile output. Bakk, Oberski, and Vermunt (in press) showed that if a continuous distal outcome has error variances which differ across latent classes, also a quadratic effect should be included in the model for the classes.

Aside from a ML or BCH adjusted step-three analysis, one may perform an unadjusted step-three analysis. For modal class assignment, this is equivalent to using the modal class assignments outside Latent GOLD. For proportional class assignment, it is similar to what Latent GOLD does when it provides ProbMeans and Profile information for inactive covariates.

6.2 Obtaining the Score Equation

The posterior class membership probability $P(x|\mathbf{z}_i, \mathbf{y}_i)$ can be written as a logistic function of the $\mathbf{z}_i$ and $\mathbf{y}_i$ variables included in the estimated LC model. That is,

$$P(x|\mathbf{z}_i, \mathbf{y}_i) = \frac{\exp(\zeta_{0x} + \sum_{p=1}^{P} \zeta_{px} \cdot z_{ip} + \sum_{t=1}^{T} \zeta_{P+t,x} \cdot y_{it})}{\sum_{x'=1}^{K} \exp(\zeta_{0x'} + \sum_{p=1}^{P} \zeta_{px'} \cdot z_{ip} + \sum_{t=1}^{T} \zeta_{P+t,x'} \cdot y_{it})}.$$

This is what what we refer to as the scoring in equation. Note that for nominal z_i and y_{it}, the appropriate contrasts (dummies or effects) should be included in the scoring equation.

The Step3-Scoring option computes the ζ coefficients of the above scoring equation. The computation is, in fact, the same as of an unadjusted step-three analysis using proportional class assignment. The major difference is that rather than relating class membership to new variables, it is related to the same variables used in the original (step-1) model.

The scoring equation obtained in this way will give a perfect representation of the posterior membership probabilities when the error (co)variances of continuous indicators and the direct effects between categorical indicators are class independent and when there are no missing values on the indicators. However, when the original LC model contains continuous indicators with class-dependent error variances also *quadratic terms* $(y_{it})^2$ should be included for the indicators concerned, and when the model contains class-dependent covariances or direct associations (local dependencies), also the *interaction terms* $(y_{it} \cdot y_{it'})$ should be included for these indicators.

Moreover, when the original model is run with the option to retain cases with missing values and the scoring equation is also computed with missing values included, one should include *missing value dummies* in the scoring equation. This yields separate ζ coefficients for the missing value categories of the indicators concerned. A missing value dummy is a variable which is added to the scoring equation, taking the value 1 if the variable it refers to is missing and 0 otherwise. It should be noted that the missing values themselves are handled in the standard way; that is, they are imputed with either the mean or the average effect, depending on whether the variable is numeric or nominal (see section 7.2.2). The advantage of the Missing Value Dummies option (rather than excluding missing values) is that when scoring new observations, missing values will automatically be handled appropriately.

6.3 Estimation of Step-Three Models and the Scoring Equation

The Latent GOLD 5.1 implementation of the estimation of step-three models and the scoring equation is as follows:

1. The data set is expanded to contain one record for each (assigned) latent class for every case. Moreover, a variable indicating the assigned class number is appended to this expanded data file.

2. For the ML and BCH adjustment procedures, the rescaled classification table entries $P(a_i|x)$ are computed.

3. The "case" weights for the expanded data set are computed. For modal ML (or none), the weights equal 1 for the modal class and 0 for the other classes. For proportional ML (or none), the weights equal the posterior membership probabilities. For BCH, the weights are obtained by multiplying the "ML weights" by the inverse of the rescaled classification table (see Vermunt, 2010).

4. The fixed values for $P(a_i|x)$ are specified. For ML adjustment, these are the numbers computed in step 2. For BCH and no adjustment, these are 1 for the diagonal entries ($a_i = x$) and 0 for the other entries, which corresponds to specifying that there are no classification errors. Note that in the BCH approach it is the case weights that account for

the classification errors. In step-three dependent models, also the fixed values for $P(x)$ are specified.

5. One of the LC models defined in equations (13) and (14) is estimated using the standard Latent GOLD estimation procedures. In the dependent option, a separate model is estimated for each dependent variable. No random starting values are used, but instead covariate effects get starting values of 0 and dependent variables get starting values corresponding to their overall proportions/means. Robust standard errors are used to account for the fact that the model is estimated with multiple weighted records per case (see Vermunt, 2010).

The scoring equation is estimated using proportional class assignment without adjustment for classification errors.

7 Estimation and Other Technical Issues

7.1 Log-likelihood and Log-posterior Function

The parameters of the various types of LC models are estimated by means of Maximum Likelihood (ML) or Posterior Mode (PM) methods. The likelihood function is derived from the probability density function defined in equation (1). Let $\boldsymbol{\vartheta}$ denote the vector containing the unknown γ and β parameters. As before, $\mathbf{y}_i$ and $\mathbf{z}_i$ denote the vectors of dependent and explanatory variables for case i, and I denotes the total number of cases.

ML estimation involves finding the estimates for $\boldsymbol{\vartheta}$ that maximize the log-likelihood function

$$\log \mathcal{L} = \sum_{i=1}^{I} w_i \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta}).$$

Here, $f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})$ is the probability density associated with case i given parameter values $\boldsymbol{\vartheta}$ and w_i is the *Case Weight* corresponding to case i.[22] This case weight w_i can be used to group identical response patterns or to specify (complex survey) sampling weights. In the former case, w_i will serve as a frequency count, and in the latter case, Latent GOLD will provide pseudo

[22] In order to simplify the discussion, in this section we discuss only on the situation without known-class indicators.

ML estimates (Patterson, Dayton, and Graubard, 2002).[23] The other type of weight – *Replication Weight* v_{it} – that was introduced in the previous section modifies the definition of the relevant probability density $f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})$.

In order to prevent boundary solutions or, equivalently, to circumvent the problem of non-existence of ML estimates, we implemented some ideas from Bayesian statistics in Latent GOLD. The boundary problems are that multinomial probabilities and Poisson rates may become zero, and that error variances in normal models may converge to zero. The first problem is circumvented by using Dirichlet priors for the latent and the conditional response probabilities and gamma priors for the Poisson rates, and the second by using inverse-Wishart priors for the error variance-covariance matrices (Clogg et al., 1991; Galindo-Garre, Vermunt, and Bergsma, 2004: Gelman et. al., 1996; Schafer, 1997). These are so-called conjugate priors since they have the same form as the corresponding multinomial, Poisson, and multivariate normal probability densities. The implication of using priors is that the estimation method is no longer ML but Posterior Mode (PM) estimation.

Denoting the assumed priors for $\boldsymbol{\vartheta}$ by $p(\boldsymbol{\vartheta})$ and the posterior by $\mathcal{P}$, PM estimation involves finding the estimates for $\boldsymbol{\vartheta}$ that maximize the log-posterior function

$$\begin{aligned} \log \mathcal{P} &= \log \mathcal{L} + \log p(\boldsymbol{\vartheta}) \\ &= \sum_{i=1}^{I} w_i \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta}) + \log p(\boldsymbol{\vartheta}), \end{aligned}$$

or, in other words, finding the point where $\frac{\partial \log \mathcal{P}}{\partial \boldsymbol{\vartheta}} = 0$. The algorithms used to solve this problem are described below.

The user-defined parameters in the priors $p(\boldsymbol{\vartheta})$ can be chosen in such a way that $\log p(\boldsymbol{\vartheta}) = 0$, which makes PM estimation turn into ML estimation. PM estimation can also be seen as a form of penalized ML estimation, in which $p(\boldsymbol{\vartheta})$ serves as a function penalizing solutions that are too near to the boundary of the parameter space and, therefore, smoothing the estimates away from the boundary.

[23] In Latent GOLD Advanced, there is a more elegant option for dealing with sampling weights, as well as with other complex survey sampling features.

7.2 Missing Data

7.2.1 Indicators and dependent variable

Latent GOLD provides full information ML or PM estimates with missing data in the response variables. Let us look at a simple example in order to get an impression on how this works. Suppose that we have a LC Cluster model with four locally independent indicators. The log-likelihood contribution of case i for which we have observed all four variables observed is, of course, equal to

$$\log \mathcal{L}_i = \log \sum_{x=1}^{K} P(x)\, f(y_{i1}|x)\, f(y_{i2}|x)\, f(y_{i3}|x)\, f(y_{i4}|x).$$

On the other hand, the contribution of another case i' for which the value of the second indicator is missing equals

$$\log \mathcal{L}_{i'} = \log \sum_{x=1}^{K} P(x)\, f(y_{i'1}|x)\, f(y_{i'3}|x)\, f(y_{i'4}|x).$$

As can be seen, the likelihood contribution is based on the observed indicators only, which means that for case i', in fact, we have a model with only three instead of four indicators. It shows that parameters are estimated using all available information for each of the cases. The assumption that is made is that the missing data are missing at random (MAR) or, equivalently, that the missing data mechanism is ignorable (Little and Rubin, 1987; Schafer, 1997; Skrondal and Rabe-Hesketh, 2004; Vermunt, 1997).

Although conceptually similar, technically things are somewhat more complicated with multivariate densities for multiple response variables, which are needed with local dependencies. Let us take the same example with four indicators, but now assume that indicators one, two, and three are correlated within latent classes. The likelihood contribution for the same persons i and i' would now be

$$\begin{aligned} \log \mathcal{L}_i &= \log \sum_{x=1}^{K} P(x)\, f(y_{i1}, y_{i2}, y_{i3}|x)\, f(y_{i4}|x), \\ \log \mathcal{L}_{i'} &= \log \sum_{x=1}^{K} P(x)\, f(y_{i'1}, y_{i'3}|x)\, f(y_{i'4}|x). \end{aligned}$$

As can be seen, for case i' who has a missing value in the set containing indicators one, two, and three, $f(y_{i'1}, y_{i'2}, y_{i'3}|x)$ is replaced by $f(y_{i'1}, y_{i'3}|x)$,

which is obtained by summing over values of the variable with missing values; that is, $f(y_{i'1}, y_{i'3}|x) = \sum_{m=1}^{M} f(y_{i'1}, y_{i'2} = m, y_{i'3}|x)$. In other words, the multivariate density $f(\mathbf{y}_{ih}|x, \mathbf{z}_i)$ corresponding to set h is based on the observed variables for the person concerned, and is obtained by summing over all possible values of the variables with missing values.

In the case of missing data, it is important to clarify the interpretation of the chi-squared goodness-of-fit statistics. Although parameter estimation with missing data is based on the MAR assumption, the chi-squared statistics not only test whether the model of interest holds, but also the much more restrictive MCAR (missing completely at random) assumption (see Vermunt, 1997). Thus, caution should be used when interpreting the overall goodness-of-fit tests in situations in which there is missing data.[24]

7.2.2 Covariates and predictors

If the technical option for including missing values on covariates and predictors is off, cases with missing covariate values and replications with missing predictor values are excluded from the analysis. When this technical option is on, such cases and replications are retained by imputing the missing values using the methods described below.

Missing values on numeric predictors and covariates are replaced by the sample mean. This is the mean over all cases without a missing value for covariates and the mean over all replications without a missing value for predictors. Missing values on nominal predictors and covariates are handled directly by the design matrix. In fact, the effect for the missing value category is equated to the average effect for the categories. Recall the effect coding scheme illustrated in subsection 2.4 for the case of a nominal predictor with 4 categories. Suppose there is also a missing category. The design matrix

[24] For Syntax models, Latent GOLD has an option (MARchi2) to obtain adjusted chi-squared statistics under the MAR assumption. By subtracting the chi-squared values of the saturated model from those of the Latent GOLD model of interest and adjusting the number of degrees of freedom in the appropriate manner, we obtain the corresponding chi-squared tests under MAR.

that is set up for the 3 non-redundant terms is then

category 1	1	0	0
category 2	0	1	0
category 3	0	0	1
category 4	−1	−1	−1
missing	0	0	0

.

As can be seen, the entries corresponding to the missing category are all equal to 0, which amounts to setting its coefficient equal to zero. Since in effect coding the unweighted mean of the coefficients equals zero, equating the effect of the missing value category to zero implies that it is equated to the unweighted average of the other four categories. This imputation method for nominal variables is therefore similar to mean imputation with numeric variables.

In the case of dummy coding with the first category as the reference category, the design matrix that is set up for the 3 non-redundant terms is

category 1	0	0	0
category 2	1	0	0
category 3	0	1	0
category 4	0	0	1
missing	1/4	1/4	1/4

.

The number 1/4 (one divided by the number of categories of the nominal variable concerned) implies that the parameter of the missing value category is equated to the unweighted mean of the parameters of the other four categories. Note that the coefficient for the reference category is fixed to 0. Also with "dummy last", we would get a row with 1/4s for the missing value category.

7.2.3 Summary of the Missing Value Settings

The Latent GOLD Missing Values option allows for the inclusion of records containing missing values on covariates and predictors (as well as records containing missing values on the indicators). The solution we have chosen for covariates and predictors with missing values is to impute the mean for the scale type numeric and equate the effect of the missing value category to zero for the scale type nominal. Missing values on indicators and dependent variables are handled directly in the likelihood function.

Exclude cases: For Regression, selection of this option excludes all replications having missing values on the dependent variable or any of the predictors and all cases having missing values on any of the active covariates. For Cluster/DFactor, selection of this option excludes all cases having missing values on any of the indicators or active covariates.

Include indicators/dependent: For Regression, selection of this option excludes replications having missing values on any of the predictors and cases having missing values on any of the active covariates. For Cluster/DFactor, selection of this option excludes cases having missing values on any of the active covariates. After exclusion of these cases, the remaining cases with missing values on the dependent variable (Regression) or on any indicator (Cluster and DFactor) are included in the analysis and handled directly in the likelihood function.

Include all: Selection of this option includes all cases and replications in the analysis regardless of the presence of missing values. Cases or replications with missing values on the dependent variable (Regression) or on any indicator (Cluster and DFactor) are included in the analysis and handled directly in the likelihood function. Missing values on Predictors (Regression submodule), or active covariates (Regression, Cluster and DFactor submodules) are imputed using Latent GOLD's imputation procedure. Inclusion in a model of covariates designated as inactive has no effect on which cases are excluded. Therefore, these missing values options have no effect with respect to the presence or absence of missing values on covariates specified to be inactive.

7.3 Prior Distributions

The different types of priors have in common that their user-defined parameters (*Bayes Constants*) denoted by α can be interpreted as adding α observations – for instance, the program default of one – generated from a conservative null model (as is described below) to the data. All priors are defined in such a way that if the corresponding α's are set equal to zero, $\log p(\boldsymbol{\vartheta}) = 0$, in which case we will obtain ML estimates. Below we present the relevant $\log p(\boldsymbol{\vartheta})$ terms without their normalizing constants.

Let the symbol U_h denote the number of unique covariate and predictor patterns in the model for subset h, and u_h a particular unique pattern. We use $h = 0$ to refer to the model for the latent variable(s). In Regression models there is only one subset; that is, $h = 1$ refers to the model for the

dependent variable. In Cluster and DFactor models, $1 \leq h \leq H$, where H is the number of locally independent subsets of indicators.

The Dirichlet prior for the latent class probabilities equals

$$\log p\left[\pi_{x|\mathbf{z}_{u_o}}\right] = \frac{\alpha_1}{K \cdot U_0} \log \pi_{x|\mathbf{z}_{u_o}}.$$

Here, K denotes the number of latent classes and α_1 the *Bayes Constant* to be specified by the user. As can be seen, the influence of the prior is equivalent to adding $\frac{\alpha_1}{K}$ cases to each latent class. These cases are distributed evenly over the various covariate patterns. This prior makes the sizes of the latent classes slightly more equal and the covariate effects somewhat smaller.

For categorical response variables, we use the following Dirichlet prior:

$$\log p\left[\pi_{\mathbf{m}|h,x,\mathbf{z}_{u_h}}\right] = \frac{\widehat{\pi}_{\mathbf{m}|h}\,\alpha_2}{K \cdot U_h} \log \pi_{\mathbf{m}|h,x,\mathbf{z}_{u_h}}.$$

where the $\widehat{\pi}_{\mathbf{m}|h}$ are based on the observed marginal distributions of the response variables in set h. In the univariate case with a single variable in set h, $\widehat{\pi}_{\mathbf{m}|h}$ is simply the observed marginal distribution $\widehat{\pi}_{m|h}$. In the multivariate case, $\widehat{\pi}_{\mathbf{m}|h}$ are estimated probabilities in the independence model; that is, $\widehat{\pi}_{\mathbf{m}|h} = \prod_{t \in h} \widehat{\pi}_{m_t|t}$. This Dirichlet prior can be interpreted as adding $\frac{\alpha_2}{K}$ observations to each latent class with preservation of the observed marginal item distributions, where α_2 is a parameter to be specified by the user. The $\frac{\alpha_2}{K}$ observations are distributed evenly over the observed covariate/predictor patterns. This prior makes the class-specific response probabilities slightly more similar to each other and smooths the β parameters somewhat towards zero.

For binomial counts, we use the same Dirichlet prior as for categorical variables:

$$\log p\left[\pi_{h,x,\mathbf{z}_{u_h}}\right] = \frac{\widehat{\pi}_h\,\alpha_2}{K \cdot U_h} \log \pi_{h,x,\mathbf{z}_{u_h}}.$$

where $\widehat{\pi}_h$ is the overall observed binomial probability.

For Poisson counts, we implemented a gamma prior. Let $\widehat{\theta}_h$ be the overall observed Poisson rate. The prior we use has the form

$$\log p\left[\theta_{h,x,\mathbf{z}_{u_h}}\right] = \frac{\alpha_3}{K} \log\left[\frac{\theta_{h,x,\mathbf{z}_{u_h}}\alpha_3}{K \cdot \widehat{\theta}_h}\right] - \frac{\theta_{h,x,\mathbf{z}_{u_h}}\alpha_3}{K \cdot \widehat{\theta}_h}.$$

This prior can be interpreted as adding $\frac{\alpha_3}{K}$ events to each latent class with preservation of the overall Poisson rate $\widehat{\theta}_h$.

The inverse-Wishart priors[25] for the error variance-covariance matrices are of the form

$$\log p(\boldsymbol{\Sigma}_{h|x}) = -0.5\frac{\alpha_4}{K}\log\left|\boldsymbol{\Sigma}_{h|x}\right| - 0.5\frac{\alpha_4}{K}\text{tr}\left(\mathbf{D}_{s_h^2}\boldsymbol{\Sigma}_{h|x}^{-1}\right).$$

Here, $\mathbf{D}_{\mathbf{s}_h^2}$ is a diagonal matrix containing the observed variances of the K_h dependent variables belonging to set h. This prior can be interpreted as incrementing each latent class with $\frac{\alpha_4}{K}$ observations which are at a distance of one standard deviation of the class-specific mean and which have covariances of zero. Here, α_4 is the parameter to be specified by the user. This prior slightly increases the class-specific error variances and slightly decreases the class-specific covariances.

The influence of the priors on the final parameter estimates depends on the values chosen for the α's, as well as on the sample size. The default settings are $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1.0$. This means that with moderate sample sizes the influence of the priors on the parameter estimates is negligible. Setting $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ yields ML estimates.[26]

7.4 Algorithms

To find the ML or PM estimates for the model parameters $\boldsymbol{\vartheta}$, Latent GOLD uses both the EM and the Newton-Raphson algorithm. In practice, the estimation process starts with a number of EM iterations. When close enough to the final solution, the program switches to Newton-Raphson. This is a way to exploit the advantages of both algorithms; that is, the stability of EM even when it is far away from the optimum and the speed of Newton-Raphson when it is close to the optimum.

[25] Actually, our prior differs somewhat from a real inverse-Wishart distribution since we omit the term $-\frac{K_h+1}{2}\log\left|\boldsymbol{\Sigma}_{h|\mathbf{x}}\right|$, where K_h is the number of variables in set h. The reason for doing this is that this term does not become zero if $\alpha_4 = 0$, which is something we want to happen in order to be able to switch from PM to ML estimation.

[26] For Syntax models, Latent GOLD has three additional options for specifying priors: "perclass" indicates that the numbers specified represent the amount of pseudo-data added to each class (instead of the *total* amount of pseudo-data), "percategory" that the number specified with the *categorical* option represents the average amount of pseudo-data added per category of the categorical variables, and "uniform" that the number specified with the *categorical* option is divided equally across categories, rather than based on the marginal distribution of the variable concerned.

The task to be performed for obtaining PM estimates for $\boldsymbol{\vartheta}$ is finding the parameter values for which

$$\frac{\partial \log \mathcal{P}}{\partial \boldsymbol{\vartheta}} = \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\vartheta}} + \frac{\partial \log p(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} = 0. \tag{15}$$

Here,

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \boldsymbol{\vartheta}} &= \sum_{i=1}^{I} w_i \frac{\partial \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \\ &= \sum_{i=1}^{I} w_i \frac{\partial \log \sum_{x=1}^{K} P(x|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|x, \mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \\ &= \sum_{i=1}^{I} \sum_{x=1}^{K} w_{xi} \frac{\partial \log P(x|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|x, \mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}}, \end{aligned} \tag{16}$$

where

$$w_{xi} = w_i \, P(x|\mathbf{z}_i, \mathbf{y}_i, \boldsymbol{\vartheta}) = w_i \frac{P(x|\mathbf{z}_i, \boldsymbol{\vartheta}) f(\mathbf{y}_i|x, \mathbf{z}_i, \boldsymbol{\vartheta})}{f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}. \tag{17}$$

The *EM algorithm* is a general method for dealing with ML estimation with missing data (Dempster, Laird, and Rubin, 1977; McLachlan and Krishnan, 1997). This method exploits the fact that the first derivatives of the incomplete data log-likelihood ($\log \mathcal{L}$) equal the first derivatives of the complete data log-likelihood ($\log \mathcal{L}^c$). The complete data is the log-likelihood that we would have if we knew to which latent class each case belongs:

$$\begin{aligned} \log \mathcal{L}^c &= \sum_{i=1}^{I} \sum_{x=1}^{K} w_{xi} \log P(x|\mathbf{z}_i, \boldsymbol{\vartheta}) \, f(\mathbf{y}_i|x, \mathbf{z}_i, \boldsymbol{\vartheta}) \\ &= \sum_{i=1}^{I} \sum_{x=1}^{K} w_{xi} \log P(x|\mathbf{z}_i, \boldsymbol{\vartheta}) \\ &\quad + \sum_{i=1}^{I} \sum_{x=1}^{K} w_{xi} \sum_{h=1}^{H} \log f(\mathbf{y}_{ih}|x, \mathbf{z}_i, \boldsymbol{\vartheta}). \end{aligned} \tag{18}$$

Each νth cycle of the EM algorithm consist of two steps. In the Expectation (E) step, estimates $\widehat{w}_{xi}^{\nu}$ are obtained for w_{xi} via equation (17) filling in $\widehat{\boldsymbol{\vartheta}}^{\nu-1}$ as parameter values. The Maximization (M) step involves finding new $\widehat{\boldsymbol{\vartheta}}^{\nu}$ improving $\log \mathcal{L}^c$. Note that we actually use PM rather than ML

estimation, which means that in the M step we update the parameters in such a way that

$$\log \mathcal{P}^c = \log \mathcal{L}^c + \log p(\boldsymbol{\vartheta}) \tag{19}$$

increases rather than (18). Sometimes closed-form solutions are available in the M step. In other cases, standard iterative methods can be used to improve the complete data log-posterior defined in equation (19). Latent GOLD uses iterative proportional fitting (IPF) and unidimensional Newton in the M step (see Vermunt 1997, Appendices).[27]

Besides the EM algorithm, we also use a *Newton-Raphson* (NR) method.[28] In this general optimization algorithm, the parameters are updated as follows:

$$\widehat{\boldsymbol{\vartheta}}^{\nu} = \widehat{\boldsymbol{\vartheta}}^{\nu-1} - \varepsilon\ \mathbf{H}^{-1}\mathbf{g}.$$

The gradient vector $\mathbf{g}$ contains the first-order derivatives of the log-posterior to all parameters evaluated at $\widehat{\boldsymbol{\vartheta}}^{\nu-1}$, $\mathbf{H}$ is the Hessian matrix containing the second-order derivatives to all parameters, and ε is a scalar denoting the step size. Element g_k of $\mathbf{g}$ equals

$$g_k = \sum_{i=1}^{I} w_i \frac{\partial \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \vartheta_k} + \frac{\partial \log p(\boldsymbol{\vartheta})}{\partial \vartheta_k}, \tag{20}$$

and element $H_{kk'}$ of $\mathbf{H}$ equals

$$H_{kk'} = \sum_{i=1}^{I} w_i \frac{\partial^2 \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta})}{\partial \vartheta_k \partial \vartheta_{k'}} + \frac{\partial^2 \log p(\boldsymbol{\vartheta})}{\partial \vartheta_k \partial \vartheta_{k'}}. \tag{21}$$

Latent GOLD computes these derivatives analytically. The step size ε ($0 < \varepsilon \leq 1$) is needed to prevent decreases of the log-posterior to occur. More precisely, when a standard NR update $-\mathbf{H}^{-1}\mathbf{g}$ yields a decrease of the log-likelihood, the step size is reduced until this no longer occurs.

The matrix $-\mathbf{H}^{-1}$ evaluated at the final $\widehat{\boldsymbol{\vartheta}}$ yields the standard estimate for the asymptotic variance-covariance matrix of the model parameters: $\widehat{\Sigma}_{standard}(\boldsymbol{\vartheta}) = -\widehat{\mathbf{H}}^{-1}$.[29] Latent GOLD also implements two alternative estimates for $\Sigma(\boldsymbol{\vartheta})$. The first alternative is based on the outer-product of the

[27] In mixtures models for multivariate normal continuous variables, we use a Fisher-scoring algorithm within the M step when the covariance-structure parameters have no closed-form solution.

[28] Haberman (1988) proposed estimating standard LC models by Newton Raphson.

[29] The matrix $-\mathbf{H}$ is usually referred to as the observed information matrix, which serves as an approximation of the expected information matrix.

cases' contributions to the gradient vectors; that is, $\widehat{\Sigma}_{outer}(\boldsymbol{\vartheta}) = \widehat{\mathbf{B}}^{-1}$, where element $B_{kk'}$ of $\mathbf{B}$ is defined as

$$B_{kk'}=\frac{N}{N-1}\sum_{i=1}^{I} w_i \frac{\partial \log f(\mathbf{y}_i|\mathbf{z}_i,\boldsymbol{\vartheta})}{\partial \vartheta_k}\frac{\partial \log f(\mathbf{y}_i|\mathbf{z}_i,\boldsymbol{\vartheta})}{\partial \vartheta_{k'}}. \tag{22}$$

Note that $\mathbf{B}$ is the sample covariance matrix of the case-specific contributions to the elements of the gradient vector.

The third estimator for $\Sigma(\boldsymbol{\vartheta})$ is the so-called robust, sandwich, or Huber-White estimator, which is defined as

$$\widehat{\Sigma}_{robust}(\boldsymbol{\vartheta}) = \widehat{\mathbf{H}}^{-1}\,\widehat{\mathbf{B}}\;\widehat{\mathbf{H}}^{-1}.$$

The advantage of $\widehat{\Sigma}_{outer}(\boldsymbol{\vartheta})$ compared to the other two is that is much faster to compute because it uses only first derivatives. It may thus be an alternative for $\widehat{\Sigma}_{standard}(\boldsymbol{\vartheta})$ in large models. The advantage of the robust method is that contrary to the other two methods, it does not rely on the assumption that the model is correct.

Note that $\widehat{\Sigma}(\boldsymbol{\vartheta})$ can be used to obtain the standard error for any function $h(\widehat{\boldsymbol{\vartheta}})$ of $\widehat{\boldsymbol{\vartheta}}$ by the delta method:

$$\widehat{se}\left(h(\widehat{\boldsymbol{\vartheta}})\right) = \sqrt{\left(\frac{\partial h(\widehat{\boldsymbol{\vartheta}})}{\partial \widehat{\boldsymbol{\vartheta}}}\right)' \widehat{\Sigma}(\boldsymbol{\vartheta}) \left(\frac{\partial h(\widehat{\boldsymbol{\vartheta}})}{\partial \widehat{\boldsymbol{\vartheta}}}\right)}. \tag{23}$$

Latent GOLD uses the delta method, for example, to obtain standard errors of means, probabilities, and redundant parameters.

Inequality restrictions – needed for ordered Clusters, order-restricted predictor effects, and positive error variances and dispersion parameters – are dealt with using an active-set variant of the Newton-Raphson method described above (Galindo, Vermunt, Croon, 2001; Gill, Murray, and Wright, 1981). For that purpose, the effects involved in the order constraints are reparameterized so that they can be imposed using simple nonnegativity constraints of the form $\vartheta \geq 0$. In an active-set method, the equality constraint associated with an inequality constraint is activated if it is violated (here, parameter is equated to 0 if it would otherwise become negative), and inactivated if its update yields an admissible value (here, a positive update).

7.5 Convergence

The exact algorithm implemented in Latent GOLD works as follows. The program starts with EM until either the maximum number of EM iterations

(*Iteration Limits EM*) or the EM convergence criterion (*EM Tolerance*) is reached. Then, the program switches to NR iterations which stop when the maximum number of NR iterations (*Iteration Limits Newton-Raphson*) or the overall converge criterion (*Tolerance*) is reached. The convergence criterion that is used is

$$\sum_{p=1}^{npar} \left| \frac{\widehat{\vartheta}_p^{\nu} - \widehat{\vartheta}_p^{\nu-1}}{\widehat{\vartheta}_p^{\nu-1}} \right|,$$

which is the sum of the absolute relative changes in the parameters. The program also stops iterating when the change in the log-posterior is negligible, i.e., smaller than 10^{-12}.

The program reports the iteration process in the Iteration Detail output file listing. Thus, it can easily be checked whether the maximum number of iterations is reached without convergence. In addition, a warning is given if one of the elements of the gradient is larger than 10^{-3}.

It should be noted that sometimes it is more efficient to use only the EM algorithm, which is accomplished by setting *Iteration Limits Newton-Raphson*=0 in the Technical Tab. This is, for instance, the case in models with many parameters. With very large models, one may also consider suppressing the computation of standard errors and Wald statistics.

7.6 Classification EM Algorithm

As an alternative to maximum likelihood, one may choose to estimate the parameters of a latent class model by maximizing the classification log-likelihood. For this purpose, we use the classification EM algorithm (Celeux and Govaert, 1992). The procedure is similar to K-means clustering since it yields a hard-partitioning of the sample (Vermunt, 2011), and in fact, with specific restrictions, it provides a K-means solution as a special case when all variables are continuous. It can also provide many generalizations to K-means involving continuous and/or categorical scale types.

The classification log-likelihood is similar to the expected complete data log-likelihood defined in equation (18), but with the difference that the posterior $P(x|\mathbf{z}_i, \mathbf{y}_i, \boldsymbol{\vartheta})$ appearing in the definition of w_{xi} equals either 0 or 1. That is, it equals 1 if person i is assigned to class x, and 0 otherwise. The difference between classification EM and regular EM is therefore that in the E step individuals are assigned to the modal class instead of being assigned

proportionally to each class according to the current estimates of the posterior membership probabilities.

7.7 Parallel Processing

An important feature of Latent GOLD 5.1 is the use of parallel processing. This is achieved by running a model using multiple threads, where the (maximum) number of threads can be specified by the user. The Technical tab contains an option *Threads*. Here one can set the maximum number of threads to be used by Latent GOLD. The default is "all", implying that the maximum number of threads is set to equal the number of processors available on the machine.

Latent GOLD 5.1 uses two different types of parallelizations for its computations. The first involves distributing multiple estimation runs of the same model across the available processors. This is what happens in the starting values and the bootstrap procedures, where the multiple runs concern the different start sets and the bootstrap replications, respectively. This type of parallelization speeds up computations by a factor close to the number of available processors. The second type of parallelization involves performing parts of the computations of a single estimation run in parallel for portions of the data set, and subsequently combining (typically adding up) the obtained information. This is used in the E step of the EM iterations and in the computation of the derivatives for the Newton-Raphson iterations and the standard errors estimation. This will speed up the estimation in larger problems (large data sets, models with many parameters, and/or models with many latent classes), where the faster computations outweigh the additional "administration" of creating the threads and combining their results.

7.8 Start Values

Latent GOLD generates random start values. These differ every time that a model is estimated because the seed of the random number generator is obtained from the system time, as long as the technical option *Seed* equals 0. The seed used by the program is reported in the output. A run can be replicated by specifying the reported best start seed as *Seed* in the Technical Tab and setting the number of *Random Sets* to zero.

Since the EM algorithm is extremely stable, the use of random starting values is generally good enough to obtain a converged solution. However,

there is no guarantee that such a solution is also the global PM or ML solution. A well-known problem in LC analysis is the occurrence of local maxima, which also fulfill the conditions defined by likelihood equations given in (15).

The best way to prevent ending up with a local solution is the use of multiple sets of starting values since different sets of starting values may yield solutions with different log-posterior values. In Latent GOLD, the use of such multiple sets of random starting values is automated. The user can specify how many sets of starting values the program should use by changing the *Random Sets* option in the Technical Tab. Another relevant parameter is *Iterations* specifying the number of iterations to be performed per start set. More precisely, within each of the random sets, Latent GOLD performs the specified number of EM iterations. Subsequently, with the best 10 percent (rounded upwards) in terms of log-posterior, the program performs an extra 2 times *Iterations* EM iterations. Finally, it continues with the best solution until convergence. It should be noted that such a procedure increases considerably the probability of finding the global PM or ML solution, especially if both parameters are set large enough, but in general does not guarantee that it will be found in a single run.

When a model contains two or more latent classes or one or more DFactors, the starting values procedure will generate the specified number of starting sets and perform the specified number of iterations per set. In one-class models in which local maxima may occur – for example, in models with continuous factors (see Advanced option) – both the specified number of starting sets and iterations per set are reduced by a factor of three. In one-class models in which local maxima cannot occur, the number of starting sets is automatically equated to 1.

With the option *Tolerance*, one can specify the EM convergence criterion to be used within the random start values procedure. Thus, start values iterations stop if either this tolerance or the maximum number of iterations is reached.

7.9 Bootstrapping the P Value for Chi^2 Statistics or -2LL Difference

Rather than relying on the asymptotic p value, it also possible to estimate the p value associated with the goodness-of-fit chi-squared statistics by means of

a parametric bootstrap. This option is especially useful with sparse tables (Langeheine, Pannekoek, and Van de Pol, 1996) and with models containing order restrictions (Galindo and Vermunt, 2005; Vermunt, 1999, 2001). Bootstrap p values will also be provided for the dissimilarity index, the total BVR, and the bivariate residuals, which are measures for which the asymptotic distribution is unknown, meaning that the p value can only be obtained by bootstrapping (Oberski, Van Kollenburg, Vermunt, 2013).

When performing a bootstrap Chi^2, the model of interest is not only estimated for the sample under investigation, but also for B replication samples. These are generated from the probability distribution defined by the ML estimates. The estimated bootstrap p value, $\widehat{p}_{boot}$, is defined as the proportion of bootstrap samples with a larger L^2 than the original sample. The Monte Carlo standard error of $\widehat{p}_{boot}$ equals $\sqrt{\frac{\widehat{p}_{boot}(1-\widehat{p}_{boot})}{B}}$. The precision of $\widehat{p}_{boot}$ can be increased by increasing the number of replications B. The number of replications is specified by the parameter *Replications*. Not only the bootstrap p values, but also the critical values (CVs) for a type-I error of 5% are reported.

A similar procedure is used to obtain a bootstrap estimate of the p value corresponding to the difference in log-likelihood value between two nested models, such as two models with different numbers of latent classes or different number of discrete factors. The -2LL-difference statistic is defined as $-2 \cdot (LL_{H_0} - LL_{H_1})$, where H_0 refers to the more restricted hypothesized model (say a K–class model) and H_1 to the more general model (say a model with $K+1$ classes). Replication samples are generated from the probability distribution defined by the ML estimates under H_0. The estimated bootstrap p value, $\widehat{p}_{boot}$, is defined as the proportion of bootstrap samples with a larger -2LL-difference value than the original sample.

The bootstrap of the -2LL-difference statistic comparing models with different numbers of latent classes was used by McLachlan and Peel (2000) in the context of mixture of normals. Nylund et al. (2007), who use the term BLRT (bootstrap likelihood-ratio test), investigated this method in the context of LC and mixture growth models. Vermunt (2001) used bootstrap p values for both the L^2 and the -2LL-difference statistic in the context of order-restricted LC models, where the L^2 measured the goodness-of-fit an ordinal LC model and the -2LL difference concerned the difference between an order-restricted and an unrestricted LC model.

The other parameter is *Seed*, which can be used to replicate a bootstrap.

The seed used by the bootstrap to generate the data sets is reported in the output.

Two technical details about the implementation of the bootstrap should be mentioned. For each bootstrap replication, the maximum likelihood estimates serve as start values. Thus, no random sets are used for the replications. Moreover, when requesting a bootstrap for the -2LL-diff, also bootstrap Chi^2 items will be reported for both models.[30]

7.10 Identification Issues

Sometimes LC models are not identified; that is, it may not be possible to obtain unique estimates for some parameters. Non-identification implies that different parameter estimates yield the same log-posterior or log-likelihood value. When a model is not identified, the observed information matrix, $-\mathbf{H}$, is not full rank, which is reported by the program. Another method to check whether a model is identified is to run the model again with different starting values. Certain model parameters are not identified if two sets of starting values yield the same $\log \mathcal{P}$ or $\log \mathcal{L}$ values with different parameter estimates.[31].

With respect to possible non-identification, it should be noted that the use of priors may make models identified that would otherwise not be identified. In such situations, the prior information is just enough to uniquely determine the parameter values.

A related problem is "weak identification", which means that even though the parameters are uniquely determined, sometimes the data is not informative enough to obtain stable parameter estimates. Weak identification can be detected from the occurrence of large asymptotic standard errors. Local solutions may also result from weak identification.

Other "identification issues" are related to the order of the latent classes of the latent variables and the uniqueness of parameters for nominal variables. For unrestricted LC Cluster and LC Regression models, the Clusters (Classes) are reordered according to their sizes: the first Cluster (Class) is always the largest Cluster (Class). For unrestricted DFactor models, the or-

[30] In Syntax models it is possible to apply the full random starting values procedure for each of the bootstrap samples.

[31] LG-Syntax contains a more sophisticated procedure to detect identification problems. It evaluates the rank of the Jacobian matrix for different random parameter values. If the rank equals the number of free model parameters, the model is (locally) identified.

dering of the DFactors is determined by the classification R-squared (largest first), and the order of the levels of each DFactor is such that the first level is larger than the last level. Parameters (γ's and β's) involving nominal variables are identified by using either effect or dummy coding, which means that parameters sum to zero over the relevant indices or that parameters corresponding to the first or last category are fixed to zero. Note that the Parameters output also contains the redundant γ and β parameters, and in the case of effect coding also their standard errors.

7.11 Selecting and Holding out Cases or Replications

7.11.1 Selecting Cases or Replications

The Select option on the Variable tab can be used to select a subset of replications/cases for the analysis. This option makes it straightforward to perform a separate analysis for different subgroups using the same data file.

When using the default select option no output is provided for the records (replications/cases) which are not used in the analysis. If one wishes such output, one should use the holdout cases or holdout replications options, rather than ignoring the unselected observations.

7.11.2 Holding out Replications

In the Regression submodule, the Select option allows specifying the unselected records to be holdout replications. This option can be used for validation purposes; that is, to determine the prediction performance of the estimated model for responses which are excluded when estimating the parameters of the specified model.

A separate set of prediction statistics is reported for the holdout replications, where for posterior and hblike prediction, the posteriors are based on the non-holdout replications. Moreover, when requesting predictions to an output file, this file will also contain the predicted values for the holdout replications.

7.11.3 Holding out Cases

Validation can also be performed by holding out cases instead of replications. Whereas parameters estimates will be based on the non-holdout cases, separate chi-squared, log-likelihood, classification, and prediction statistics are

computed for the holdout cases. Moreover, the output file will contain classification and/or prediction information for all cases, thus also for the holdout cases.

The option for holding out cases can also be used when one desires removing certain cases from the analysis, but nevertheless obtaining classification and prediction output to a file for all cases. One possible application is the analysis of very large data sets, where a subsample may be used for parameter estimation. Another application is predicting class membership for new cases based on parameter values obtained with another sample. By appending the new cases to the original data file and treating them as holdout cases, one obtains the relevant output for these cases after restoring and re-estimating the original model.

It should be noted, that classification of new cases can more easily be done using a scoring equation obtained with the new Step3-Scoring option. Moreover, in the Cluster module one can request the scoring equation directly as part of the output (option "Scoring Equations").

8 Latent GOLD's Output

Below, we provide the necessary technical details on the quantities presented in the various Latent GOLD output sections (Model Summary, Parameters, Profile, ProbMeans, Iteration Detail, Frequencies, Bivariate Residuals, Standard Classification, Model Classification, and Estimated Values), as well as on the output that can be written to files (Standard Classification, Model Classification, Predicted values, Individual Coefficients, Cook's D, and Variance-Covariance Matrix).

8.1 Model Summary

This first part of the Model Summary output section reports the number of cases ($N = \sum_{i=1}^{I} w_i$), the total number of replications (in Regression models with a case ID variable, where $N_{rep} = \sum_{i=1}^{I} w_i \sum_{t=1}^{T_i} v_{it}$), the number of estimated parameters ($npar$), the number of activated constraints (in models with order restrictions and models with normally distributed response variables), the seed used by the pseudo random number generator, the seed of the best start set, and the seed used by the bootstrap procedure.

The last part (Variable Detail) contains information on the variables used

in the analysis. The other four parts - Chi-squared Statistics, Log-likelihood Statistics, Classification Statistics, Model Classification Statistics, and Prediction Statistics - are described in more detail below.

8.1.1 Chi-squared statistics

The program reports chi-squared and related statistics, except when the data file contains replication weights unequal to 0 or 1 or when there are continuous response variables. The three reported chi-squared measures are the likelihood-ratio chi-squared statistic L^2, the Pearson chi-squared statistic X^2, and the Cressie-Read chi-squared statistic CR^2. Before giving the definitions of the chi-squared statistics, we have to explain two types of groupings that have to be performed with the original cases.

The first is the grouping of identical cases; that is, cases that have the same covariate, known-class, predictor, and exposure values, and give the same responses. This yield I^* unique data patterns with observed frequency counts denoted by n_{i^*}, where i^* denotes a particular data pattern. These frequency counts are obtained by summing the case weights w_i of the cases with data pattern i^*; that is, $n_{i^*} = \sum_{i \in i^*} w_i$.[32] In order to obtain the chi-squared statistics, we also have to group cases with identical covariate, known-class, predictor, and exposure values, which amounts to grouping cases without taking into account their responses.[33] This yields the sample sizes N_u for the U relevant multinomials, where u denotes a particular multinomial or "covariate" pattern. These sample sizes are obtained by $N_u = \sum_{i \in u} w_i$ or $N_u = \sum_{i^* \in u} n_{i^*}$.[34] Note that $N = \sum_{u=1}^{U} N_u$.

Let $\widehat{m}_{i^*}$ denote the estimated cell count for data pattern i^*, which is obtained by:

$$\widehat{m}_{i^*} = N_{u_{i^*}} \widehat{f}(\mathbf{y}_{i^*} | \mathbf{z}_{i^*}), \tag{24}$$

i.e., by the product of the total number of cases with the same "covariate" pattern as data pattern i^* ($N_{u_{i^*}}$) and the estimated multinomial probability

[32] With the somewhat loose but rather simple notation $i \in i^*$ we mean "all the cases with data pattern i^*".

[33] With missing values on some indicators, also the missing value pattern is used as a grouping criterion. That is, cases belonging to the same "covariate" pattern should also have observed values on the same set of indicators or, equivalently, the same missing data pattern.

[34] With $i \in u$ we mean "all the cases with covariate pattern u", and with $i^* \in u$ "all the data patterns with covariate pattern u".

corresponding to data pattern i^*.[35]

Using these definitions of $\widehat{m}_{i^*}$, n_{i^*}, and N, the chi-squared statistics are calculated as follows:[36]

$$\begin{aligned} L^2 &= 2\sum_{i^*=1}^{I^*} n_{i^*} \log \frac{n_{i^*}}{\widehat{m}_{i^*}}, \\ X^2 &= \sum_{i^*=1}^{I^*} \frac{(n_{i^*})^2}{\widehat{m}_{i^*}} - N, \\ CR^2 &= 1.8\sum_{i^*=1}^{I^*} n_{i^*} \left[\left(\frac{n_{i^*}}{\widehat{m}_{i^*}}\right)^{2/3} - 1\right]. \end{aligned}$$

The number of degrees of freedom is defined by

$$df = \min\left\{\sum_{u=1}^{U}\left(\prod_{t=1}^{T_u^*} M_{ut}^* - 1\right), N\right\} - npar.$$

Here, T_u^* is the total number of observed indicators (replications) in "covariate" pattern u, and M_{ut}^* denotes the number of categories of the tth observed indicator (replication) corresponding to "covariate" pattern u.[37] The term $\min\{\cdot\}$ indicates that df is based on the sample size N when the number of independent cells in the hypothetical frequency table is larger than the sample size. The chi-squared values with the corresponding df yield the asymptotic p -values, which can be used to determine whether the specified model fits the data.

If the *Bootstrap Chi*^{2} option is used, the program also provides the estimated bootstrap p-values corresponding to the X^2, L^2, and CR^2 goodness-of-fit statistics, as well as the 5% critical values (CV) and the Monte Carlo standard errors of the p-values. This option is especially useful with sparse tables, in which case the asymptotic p-values can not be trusted. A good indication of sparseness is when X^2 and L^2 take on very different values.

[35] In order to get meaningful chi-squared statistics, in models with a known-class indicator we, in addition, divide by $\sum_{x=1}^{K} \tau_{i^*x} P(x|\mathbf{z}_{i^*})$.

[36] Note that we are using a somewhat unconventional formula for X^2. The reason for this is that the sum $\sum_{i^*=1}^{I^*}$ is over the nonzero observed cells only.

[37] A binomial count can take on a values between 0 and the total exposure E_{ut}. M_{ut} is therefore equal to $E_{ut}+1$. A Poisson count can take on any value, which means that the number of categories is in fact infinity. However, we set $M_{ut} = \max(y_{ut}) + 2$, which amounts to treating all values larger than the largest observed score as a single category.

The program reports the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC), Akaike Information Criterion 3 ($AIC3$), the Consistent Akaike Information Criterion ($CAIC$), and the sample size adjusted BIC ($SABIC$) based on the L^2 and df, which is the more common formulation in the analysis of frequency tables. They are defined as

$$\begin{aligned} BIC_{L^2} &= L^2 - \log(N)\, df, \\ AIC_{L^2} &= L^2 - 2\ df, \\ AIC3_{L^2} &= L^2 - 3\ df, \\ CAIC_{L^2} &= L^2 - (\log(N) + 1)\ \ df, \\ SABIC_{L^2} &= L^2 - \log((N+2)/24)\ df. \end{aligned}$$

These information criteria weight the fit and the parsimony of a model: the lower BIC, AIC, $AIC3$, $CAIC$, or $SABIC$ the better the model.

Use of information criteria based on L^2 or $\log \mathcal{L}$ (see below) should yield the same result. The differences between BIC, AIC, $AIC3$, $CAIC$, and $SABIC$ values across models are the same with both methods. However, with extremely large df, the L^2 based information measures may become more highly negative than the maximum precision can indicate, which makes their rounded values meaningless. In such cases, one has to use the (equivalent) $\log \mathcal{L}$ based measures.

Another statistic that is provided in the chi-squared statistics section is the Dissimilarity Index (DI), which is a descriptive measure that is defined as follows:

$$DI = \frac{\left\{ \left(\sum_{i^*=1}^{I^*} |n_{i^*} - \widehat{m}_{i^*}| \right) + \left(N - \sum_{i^*=1}^{I^*} \widehat{m}_{i^*} \right) \right\}}{2\, N}.$$

It should be noted that the term $(N - \sum_{i^*=1}^{I^*} \widehat{m}_{i^*})$ captures the contribution of the zero observed cells to DI. This term has to be added to the formula because $\sum_{i^*=1}^{I^*} |n_{i^*} - \widehat{m}_{i^*}|$ is a sum over the non-zero cell counts only. DI is a descriptive measure indicating how much observed and estimated cell frequencies differ from one another; that is, indicating which proportion of the sample should be moved to another cell to get a perfect fit.

The last statistic that is provided in the chi-squared statistics section is the *Total BVR*; that is, the sum of the bivariate residuals (see subsection 8.6 for details on how the BVRs are computed). This statistic is provided only if bivariate residuals are requested.

If the *Bootstrap Chi*^{2} option is used, the program also provides the estimated bootstrap p-values for DI and *Total BVR*, which are statistics for which the asymptotic distribution is unknown. Also the 5% critical values (CV) and the Monte Carlo standard errors of the p-values are reported.

8.1.2 Log-likelihood statistics

Furthermore, we report the values of the log-likelihood ($\log \mathcal{L}$), the log-prior ($\log p(\boldsymbol{\vartheta})$), and log-posterior ($\log \mathcal{P}$) . Recall that

$$\begin{aligned} \log \mathcal{L} &= \sum_{i=1}^{I} w_i \log \widehat{f}(\mathbf{y}_i|\mathbf{z}_i), \\ \log \mathcal{P} &= \log \mathcal{L} + \log p(\widehat{\boldsymbol{\vartheta}}). \end{aligned}$$

In addition, the Bayesian Information Criterion (BIC), the Akaike Information Criterion (AIC), the Akaike Information Criterion 3 ($AIC3$),[38] the Consistent Akaike Information Criterion ($CAIC$), and the sample size adjusted BIC ($SABIC$) based on the log-likelihood are reported. These are defined as

$$\begin{aligned} BIC_{\log \mathcal{L}} &= -2 \log \mathcal{L} + \log(N)\, npar, \\ AIC_{\log \mathcal{L}} &= -2 \log \mathcal{L} + 2\ npar, \\ AIC3_{\log \mathcal{L}} &= -2 \log \mathcal{L} + 3\ npar, \\ CAIC_{\log \mathcal{L}} &= -2 \log \mathcal{L} + (\log(N) + 1)\ npar, \\ SABIC_{\log \mathcal{L}} &= -2 \log \mathcal{L} + \log((N+2)/24)\ npar. \end{aligned}$$

If the *Bootstrap -2LL diff* option is used, the program also provides the estimated bootstrap p-value (and the standard error) for the -2LL difference test between a restricted and an unrestricted model.

8.1.3 Classification statistics

This set of statistics contains information on how well we can predict to which latent class cases belong given their observed y and z values, or, in other words, how well the latent classes are separated. Classification is based

[38] Results from Andrews and Currim (2003) and Dias (2004) suggest that AIC3 is a better criterion than BIC and AIC in determining the number of latent classes in LC and FM models.

on the latent classification or posterior class membership probabilities. For response pattern i, these are calculated as follows:

$$\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i) = \frac{\widehat{P}(x|\mathbf{z}_i)\widehat{f}(\mathbf{y}_i|x, \mathbf{z}_i)}{\widehat{f}(\mathbf{y}_i|\mathbf{z}_i)}. \tag{25}$$

The numerator and denominator are the maximum likelihood estimates for the terms appearing in the general mixture model defined in equation (1).

These quantities are used to compute the estimated proportion of classifications errors (E), as well as three R^2-type measures for nominal variables: the proportional reduction of classification errors $R^2_{x,errors}$, a measure based on entropy labelled $R^2_{x,entropy}$, and a measure based on qualitative variance labelled $R^2_{x,variance}$. The former is similar to the association measure Lambda and the latter to the Goodman and Kruskal tau-b association coefficient for nominal dependent variables (Magidson, 1981).

The proportion of classification errors is defined as:

$$E = \frac{\sum_{i=1}^{I} w_i \left[1 - \max \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\right]}{N}.$$

Each of the three R^2_x measures is based on the same type of reduction of error structure; namely,

$$R^2_x = \frac{\text{Error}(x) - \text{Error}(x|\mathbf{z}, \mathbf{y})}{\text{Error}(x)}, \tag{26}$$

where $\text{Error}(x)$ is the total error when predicting x without using information on $\mathbf{z}$ and $\mathbf{y}$, and $\text{Error}(x|\mathbf{z}, \mathbf{y})$ is the prediction error if we use all observed information from the cases. $\text{Error}(x|\mathbf{z}, \mathbf{y})$ is defined as the (weighted) average of the case-specific errors $\text{Error}(x|\mathbf{z}_i, \mathbf{y}_i)$,

$$\text{Error}(x|\mathbf{z}, \mathbf{y}) = \frac{\sum_{i=1}^{I} w_i \text{Error}(x|\mathbf{z}_i, \mathbf{y}_i)}{N}.$$

The three R^2_x measures differ in the definition of $\text{Error}(x|\mathbf{z}_i, \mathbf{y}_i)$. In $R^2_{x,errors}$, it equals $1 - \max \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$, in $R^2_{x,entropy}$, $\sum_{x=1}^{K} - \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i) \log \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$, and in $R^2_{x,variance}$, $1 - \sum_{x=1}^{K} [\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)]^2$. In the computation of the total error $\text{Error}(x)$, the $\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$ are replaced by the estimated marginal latent probabilities $\widehat{P}(x)$, which are defined as

$$\widehat{P}(x) = \frac{\sum_{i=1}^{I} w_i \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)}{N} = \frac{\sum_{i=1}^{I} \widehat{w}_{xi}}{N}. \tag{27}$$

In DFactor models, we use the same definitions, but then for each of the L discrete factors x_ℓ. An exception is the definition of $R^2_{x_\ell,variance}$ for which we can use the category scores, yielding a standard variance. The contribution of case i to Error$(x_\ell|\mathbf{z}_i,\mathbf{y}_i)$ equals $\sum_{x_\ell=1}^{K_\ell}\left[x_{x_\ell}^{\ell*}-\widehat{E}(x^{\ell*})\right]^2\widehat{P}(x_\ell|\mathbf{z}_i,\mathbf{y}_i)$, where $\widehat{E}(x^{\ell*})=\sum_{x_\ell=1}^{K_\ell}x_{x_\ell}^{\ell*}\,\widehat{P}(x)$, and Error$(x_\ell)$ equals $\sum_{x_\ell=1}^{K_\ell}\left[x_{x_\ell}^{\ell*}-\widehat{E}(x^{\ell*})\right]^2\widehat{P}(x_\ell)$.

Latent GOLD reports five other classification statistics: the Classification Log-likelihood (CL), the *Entropy*, the Classification Likelihood Criterion (CLC), the Approximate Weight of Evidence (AWE), and a version of the Integrated Classification Likelihood called $ICL\text{-}BIC$. CL and *Entropy* are quantities needed to compute the other three. CLC indicates how well a model performs in terms of fit and classification performance. The AWE and $ICL\text{-}BIC$ statistics adds a third dimension to the information criteria described above; they weight fit, parsimony, and the performance of the classification (Banfield and Raftery, 1993; McLachlan and Peel, 2000).

The classification log-likelihood is equivalent to the complete data log-likelihood $\log\mathcal{L}^c$, i.e.,

$$CL=\log\mathcal{L}^c=\sum_{i=1}^{I}\sum_{x=1}^{K}\widehat{w}_{xi}\log\widehat{P}(x|\mathbf{z}_i)\widehat{f}(\mathbf{y}_i|x,\mathbf{z}_i).$$

It can also be computed as

$$CL=\log\mathcal{L}-Entropy,$$

where *Entropy* is defined as follows

$$Entropy=\sum_{i=1}^{I}w_i\sum_{x=1}^{K}-\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i)\log\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i).$$

The Classification Log-likelihood Criterion equals

$$CLC=-2CL=-2\,\log\mathcal{L}+2\,Entropy.$$

AWE is defined as

$$\begin{aligned}AWE &= CLC+2\left(\frac{3}{2}+\log N\right)npar\\ &= -2\,\log\mathcal{L}+2\,Entropy+2\left(\frac{3}{2}+\log N\right)npar,\end{aligned}$$

and $ICL\text{-}BIC$ as

$$\begin{aligned} ICL\text{-}BIC &= BIC + 2\,Entropy \\ &= -2\,\log\mathcal{L} + 2\,Entropy + \log(N)npar. \end{aligned}$$

The lower AWE and/or $ICL\text{-}BIC$ value, the better a model.

The *Classification Tables* cross-tabulate the true class memberships x against the assigned (or predicted) class memberships $\hat{x}$. That is, the count in cell entry $(x, \hat{x})$ indicates the number of persons belonging to class x and being assigned to class $\hat{x}$. This number is obtained as the sum of the class x posterior membership probabilities for the cases allocated to class $\hat{x}$. It is computed as follows:

$$\sum_{i=1}^{I} w_i \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\widehat{P}(\hat{x}|\mathbf{z}_i, \mathbf{y}_i),$$

where for modal assignment $\widehat{P}(\hat{x}|\mathbf{z}_i, \mathbf{y}_i)$, equals 1 for the modal class and 0 for the other classes, and for proportional assignment, it is just the posterior membership probability for class $\hat{x}$.

The diagonal elements ($x = \hat{x}$) are the numbers of correct classifications per latent class and the off-diagonal elements ($x \neq \hat{x}$) the corresponding numbers of misclassifications. From the classification table, one can not only see how many cases are misclassified (as indicated by the proportion of classification errors E), but also detect which are the most common types of misclassifications. If a particular off-diagonal entry is large, this means that latent classes x and $\hat{x}$ are not well separated. In DFactor models a separate set of classification tables is reported for each discrete factor.

The margins of the Classification Table show the distribution of cases across classes under modal/proportional classification (column totals) and according to the model (row totals). Except for very rare situations, for modal classification, these margins will *not* be equal to one another. This illustrates the phenomenon that modal class assignments do not conserve the estimated latent class distribution. Whereas the row totals are in agreement with the estimated classes sizes,[39] the column totals show the latent class distribution that is obtained when writing the class assignments to a file using the Latent GOLD output-to-file option. For proportional classification,

[39]There may be a very small difference, which is caused by the Bayes constant for the latent classes.

row and column totals are identical and, moreover, the classification table is symmetric.

8.1.4 Model classification statistics

These statistics indicate how well one can predict class membership from an individual's covariate values, and are therefore only of interest if the estimated model contains active covariates. The measures are similar to those reported in the section "Classification Statistics"; that is, the estimated proportion of classification errors, the proportional reduction of classification errors, an entropy-based R^2 measure, and a (qualitative) variance-based R^2 measure. The difference is that now the predictions (and computations) are based on the model probabilities $\widehat{P}(x|\mathbf{z}_i)$ instead of the posterior probabilities $\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$. Whereas the total error can still be denoted as Error(x), the model prediction error in equation (26) should now be denoted as Error$(x|\mathbf{z})$ instead of Error$(x|\mathbf{z}, \mathbf{y})$.

8.1.5 Prediction statistics

In the Regression submodule, there is a prediction statistics section, which are based on the comparison between observed and predicted responses. The predicted values (probabilities) used in the computation of the prediction statistics are denoted by $\widehat{y}_{it}$ ($\widehat{P}_{m|it}$). Latent GOLD implements three types of methods for obtaining predicted values: posterior mean prediction, Hierarchical-Bayes-like prediction, and model based prediction.

In *posterior-mean (mode) or expected (modal) a posteriori prediction*, the predicted values $\widehat{y}_{it}$ ($\widehat{P}_{m|it}$) are obtained as weighted averages of the Class-specific estimates, where the posterior membership probabilities $\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$ serve as weights; that is,

$$\begin{aligned} \widehat{y}_{it} &= \sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\, \widehat{E}(y_{it}|x, \mathbf{z}_{it}^{pred}), \\ \widehat{P}_{m|it} &= \sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\, \widehat{P}(y_{it} = m|x, \mathbf{z}_{it}^{pred}). \end{aligned}$$

Note that $\widehat{E}(y_{it}|x,\, \mathbf{z}_{it}^{pred})$ equals $\widehat{\mu}_{x,\,\mathbf{z}_{it}}$ for continuous variables, $\widehat{\theta}_{x,\,\mathbf{z}_{it}} e_{it}$ for Poisson counts, $\widehat{\pi}_{x,\,\mathbf{z}_{it}}\, e_{it}$ for binomial counts, and $\sum_{m=1}^{M} y_m^*\, \widehat{P}(y_{it} = m|x,\, \mathbf{z}_{it}^{pred})$

for ordinal variables, where y_m^* is the score assigned to category m. For truncated variables, $\widehat{E}(y_{it}|x,\, \mathbf{z}_{it}^{pred})$ is replaced by $\widehat{E}(y_{it}|x,\, \mathbf{z}_{it}^{pred},\, y_{it} > 0)$, which equals $\widehat{\theta}_{x,\,\mathbf{z}_{it}} e_{it} \,/\, [1 - \widehat{P}(0|x,\, \mathbf{z}_{it}^{pred})]$ and $\pi_{x,\,\mathbf{z}_{it}}\, e_{it} \,/\, [1 - \widehat{P}(0|x,\, \mathbf{z}_{it}^{pred})]$ for Poisson and binomial counts, and $\widehat{\mu}_{x,\,\mathbf{z}_{it}} + \widehat{\sigma}_x\, \widehat{f}(y_{it}|x,\, \mathbf{z}_{it}^{pred}) \,/\, [1 - \widehat{F}(0|x,\, \mathbf{z}_{it}^{pred})]$ for continuous variables. The Class-specific predicted values for censored normal variables are defined as $\widehat{\mu}_{x,\,\mathbf{z}_{it}}[1 - \widehat{F}(0|x,\, \mathbf{z}_{it}^{pred})] + \widehat{\sigma}_x\ \widehat{f}(y_{it}|x,\, \mathbf{z}_{it}^{pred})$. These definitions show that for zero-truncated and zero-censored response variables predictions are based on estimated expected values conditional on truncation or censoring at $y_{it} = 0$ (see, for example, Long 1997).

The most natural predicted value for an ordinal, continuous, and count variable, is the individual-specific estimated expected value $\widehat{y}_{it}$. For nominal variables, this is the mode corresponding to the category with the largest $\widehat{P}_{m|it}$. As shown below, error measures for categorical variables may also be based on the M estimated probabilities instead of a single predicted value. For categorical dependent variables, we report a *prediction table* cross-classifying observed and predicted values based on the mode.

The average prediction error can be defined in various manners. We implemented measures based on squared error (MSE), minus the log-likelihood ($-MLL$), absolute error (MAE), and, for categorical variables, also the proportion of predictions errors (PPE). The various types of error measures can be denoted by the generic name Error(model). Except for nominal dependent variables, computation of MSE and MAE is straightforward:

$$\begin{aligned} MSE &= \frac{\sum_i w_i \sum_t v_{it}[y_{it} - \widehat{y}_{it}]^2}{\sum_i w_i \sum_t v_{it}} \\ MAE &= \frac{\sum_i w_i \sum_t v_{it}|y_{it} - \widehat{y}_{it}|}{\sum_i w_i \sum_t v_{it}} \end{aligned}$$

Here, w_i and v_{it} denote case and replication weights. For nominal variables, $[y_{it} - \widehat{y}_{it}]^2$ and $|y_{it} - \widehat{y}_{it}|$ are replaced by a sum over all categories: $\sum_m [I_m(y_{it}) - \widehat{P}_{m|it}]^2$ and $\sum_m |I_m(y_{it}) - \widehat{P}_{m|it}|$, where indicator variable $I_m(y_{it})$ equals 1 if $y_{it} = m$ and otherwise 0. The mean minus log-likelihood ($-MLL$) is obtained using $\widehat{y}_{it}$ as expected value in the appropriate log density function. Again, we take the average over all cases and replications. More precisely,

$$-MLL = -\,\frac{\sum_i w_i \sum_t v_{it} \log f[y_{it}|\widehat{y}_{it}]}{\sum_i w_i \sum_t v_{it}}.$$

With categorical variables, we replace $\log f[y_{it}|\widehat{y}_{it}]$ by $\sum_m I_m(y_{it}) \ln[\widehat{P}_{m|it}]$.

The general definition of the (pseudo) R_y^2 of an estimated model is the reduction of errors compared to the errors of a baseline model. More precisely,

$$R_y^2 = \frac{\text{Error(baseline)} - \text{Error(model)}}{\text{Error(baseline)}}.$$

Our baseline prediction, $\widehat{y}_0$ or $\widehat{P}_{m|0}$ is the average predicted value or response probability according to the specified model; that is,

$$\widehat{y}_0 = \frac{\sum_i w_i \sum_t v_{it} \widehat{y}_{it}}{\sum_i w_i \sum_t v_{it}} \tag{28}$$

$$\widehat{P}_{m|0} = \frac{\sum_i w_i \sum_t v_{it} \widehat{P}_{m|it}}{\sum_i w_i \sum_t v_{it}}. \tag{29}$$

Notice that in most situations, $\widehat{y}_0$ and $\widehat{P}_{m|0}$ are simply the observed sample averages, or the predicted values in the intercept-only model. This is, however, not necessarily the case when restrictions are imposed on the intercept.

There are two other prediction methods – HB-like and model Prediction. These differ from posterior mean prediction in the definition of $\widehat{y}_{it}$ and $\widehat{P}_{m|it}$. For *HB-like prediction*, the predicted values are obtained from the individual-specific posterior mean estimates of the linear terms, denoted by $\widehat{\eta}_{it}$ or $\widehat{\eta}_{m|it}$ and defined as

$$\widehat{\eta}_{it} = \sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\, \widehat{\eta}_{x,\mathbf{z}_{it}},$$

$$\widehat{\eta}_{m|it} = \sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\, \widehat{\eta}_{m|x,\mathbf{z}_{it}}.$$

These linear terms are transformed into expected values $\widehat{y}_{it}$ and response probabilities $\widehat{P}_{m|it}$ using the appropriate inverse link function. The $\widehat{y}_{it}$ and $\widehat{P}_{m|it}$ are subsequently used in the above R_y^2 formulas. Because of the similarity with prediction in Hierarchical Bayes (HB) procedures, we call this alternative method HB prediction. Note that the way we compute $\widehat{\eta}_{it}$ and $\widehat{\eta}_{m|it}$ is equivalent to computing these numbers from the individual-specific $\widehat{\beta}_{iq}$ parameters defined in equation (30).

Model-based prediction differs from posterior mean prediction in that the prior (model) class membership probabilities $\widehat{P}(x|\mathbf{z}_i)$ are used in the formulas for $\widehat{y}_{it}$ and $\widehat{P}_{m|it}$ given in equations (28) and (29) instead of the posterior

membership probabilities $\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$. Whereas posterior mean and HB-like prediction give a good indication on the within-sample prediction performance, model-based prediction gives a good indication on the out–of-sample prediction performance.

8.2 Parameters

In the Parameters output section, the program reports the estimates obtained for the β and γ parameters appearing in the linear predictors η, the estimates for error variances and covariances σ, as well as the corresponding estimated asymptotic standard errors, $\widehat{se}(\beta)$, $\widehat{se}(\gamma)$, and $\widehat{se}(\sigma)$. These standard errors are the squared roots of the diagonal elements of the estimated variance-covariance matrix $\widehat{\Sigma}(\boldsymbol{\vartheta})$. As described earlier one of three methods can be used to obtain $\widehat{\Sigma}(\boldsymbol{\vartheta})$, yielding either a standard, outer-product based, or robust standard errors and Wald statistics.

The significance of sets of parameters can be tested by means of the reported Wald statistic labeled *Wald*. In LC Regression models, we also report a Wald statistic labeled *Wald(=)*, which tests whether regression coefficients are equal between Classes (Class Independent). The Paired Comparisons output, which is nested within Parameters, provides Wald tests comparing parameters across each pair of latent classes. This allows, for example, testing in a Cluster model which Clusters are significantly different in terms of the indicators, in a Regression model which Classes have significantly different predictor effects, or in a Step3 model for which pairs of Classes the effects of the Covariates are significantly different.[40] The general formula for a Wald statistic (W^2) is

$$W^2 = \left(\mathbf{C}'\boldsymbol{\vartheta}\right)' \left(\mathbf{C}'\,\widehat{\Sigma}(\boldsymbol{\vartheta})\mathbf{C}\right)^{-1} \left(\mathbf{C}'\boldsymbol{\vartheta}\right),$$

where the tested set of linear constraints is: $\mathbf{C}'\boldsymbol{\vartheta} = \mathbf{0}$. The Wald test is a chi-squared test. Its number of degrees of freedom equals the number of constraints. Computation of standard errors and Wald statistics can be suppressed. This option may be useful in models with many parameters.

In the Regression submodule, the parameters output contains *Class-specific* $R^2_{y|x}$ values based on MSE (see also prediction statistics). In the

[40] Note that typically one may wish to account for multiple testing, for example, by applying a Bonferroni-correction which involves dividing the type I error rate by the number of tests (by the number of pairs of classes).

computation of the Class-specific errors, each observed value y_{it} is compared with the Class-specific prediction given by the model, $\widehat{y}_{itx} = \widehat{E}(y_{it}|x, \mathbf{z}_i^{pred})$. The posterior membership probabilities $\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$ quantify the contribution of case i to the error of Class x. More precisely, MSE_x is obtained as

$$MSE_x = \frac{\sum_i w_{ix} \sum_t v_{it}[y_{it} - \widehat{y}_{itx}]^2}{\sum_i w_{ix} \sum_t v_{it}},$$

where $w_{ix} = w_i \, \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$. Similarly to the overall R_y^2, in the computation of each Class-specific $R_{y|x}^2$, we use the average $\widehat{y}_{itx}$ ($\widehat{y}_{0x}$) to derive a baseline error.

In Regression, we also report the means and standard deviations of the regression coefficients. These are the typical fixed and random effects in multilevel, mixed, or random-coefficient models. The mean equals $\sum_x P(x) \, \widehat{\beta}_{xq}$, and the standard deviation $\sqrt{\sum_{x=1}^K \widehat{P}(x)(\widehat{\beta}_{xq})^2 - \left[\sum_{x=1}^K \widehat{P}(x) \, \widehat{\beta}_{xq}\right]^2}$.

The parameters output in the Cluster and DFactor submodules contains a separate $R_{y_t}^2$ value for each indicator. These are similar to explained variances in analysis of variance and item communalities in traditional factor analysis. For the scale types ordinal, continuous, and count, these are standard explained variances; that is, they are defined as the ratio of the between-class and total variation of the y_t variable concerned. For nominal indicators, the $R_{y_t}^2$ is based on the qualitative variance (Magidson, 1981), which is the sum of the category-specific variances (see classification statistics), yielding a Goodman and Kruskal tau-b measure.

The Loadings output in Cluster and DFactor and the Correlations output in DFactor contains some further "factor-analytic" output, as described in detail by Magidson and Vermunt (2003a) and Vermunt and Magidson (2005a). This factor-analytic output is obtained by a linear approximation of the logistic and log-linear regression models for the indicators. The loadings output has an interpretation similar to factor loadings in a standard factor analysis – standardized linear effects of latent variables on indicators. The correlations output provides the DFactor-DFactor and DFactor-indicator correlations.

The Error Correlations output provides the estimates of the within-class correlations between continuous indicators variables. Often, these correlations are easier to interpreted than the covariances, which are the actual model parameters.

8.3 Profile

The Profile output contains information on

- class sizes $\widehat{P}(x)$,
- class-specific response probabilities $\widehat{\pi}_{m|t,x}$ for nominal variables,
- class-specific response probabilities $\widehat{\pi}_{m|t,x}$ and means $\sum_{m=1}^{M_t} y_m^{t*} \widehat{\pi}_{m|t,x}$ for ordinal variables,
- class-specific means $\widehat{\mu}_{t,x}$ for continuous response variables,
- class-specific rates $\widehat{\theta}_{t,x}$ for Poisson counts,
- class-specific success probabilities $\widehat{\pi}_{t,x}$ for binomial counts,
- class-specific means $\widehat{E}(z_{ir}|x)$ and probabilities $\widehat{P}(z_{ir} = a|x)$ for covariates.

Except for the covariate information, these are all transformations of the β and γ parameters – or corresponding linear terms η – to a scale that makes the profiling of the Clusters (Classes) much easier. In DFactor models, this information is provided for each discrete factor x_ℓ, as well as for the *joint* latent variable x obtained by combining the L DFactors.

The first part of the Profile output contains the estimated *marginal latent probabilities* (Cluster sizes, DFactor level sizes, Class sizes). In Cluster, Regression, and one-DFactor models without covariates, these are just the model probabilities $\widehat{P}(x)$. In models with covariates or multiple DFactors, these numbers are computed by aggregating the model probabilities $\widehat{P}(x|\mathbf{z}_i)$ over covariates values and/or other DFactors as follows:

$$\begin{aligned} \widehat{P}(x) &= \sum_{i=1}^{I} \frac{w_i}{N} \widehat{P}(x|\mathbf{z}_i) \\ \widehat{P}(x_\ell) &= \sum_{i=1}^{I} \sum_{x_{\ell'} \neq x_\ell} \frac{w_i}{N} \widehat{P}(x|\mathbf{z}_i). \end{aligned}$$

The second part of the Profile output reports the *class-specific marginal means* (*probabilities*) for all indicators. In Cluster models without direct

effects of covariates on indicators and without direct indicator-indicator relations, these are simply the means (probabilities) defining the class-specific distributions. In other cases, however, these numbers should be obtained by aggregating over the other variables involved in the submodel for the response variable concerned. As an example of this aggreagrtion process, let us consider one of the more complicated cases. Suppose we have a LC Cluster model with covariates having direct effects on indicators and with a local dependency between nominal indicators y_1 and y_2. The marginal probability $\widehat{P}(y_{i1} = m|x)$ is obtained as follows:

$$\widehat{P}(y_{i1} = m_1|x) = \frac{\sum_{i=1}^{I} \sum_{m_2=1}^{M_2} \frac{w_i}{N} \widehat{P}(x|\mathbf{z}_i) \widehat{P}(y_{i1} = m_1, y_{i2} = m_2|x, \mathbf{z}_i)}{\widehat{P}(x)}.$$

As can be seen, we aggregate the joint probability of y_{i1}, y_{i2}, conditional on x, and $\mathbf{z}_i$ over $\mathbf{z}_i$ and y_{i2}, and divide the result by the marginal latent probability $\widehat{P}(x)$.

Instead of marginal means (probabilities) one can also request *partial means* (*probabilities*). These are obtained by evaluating the linear terms η at the mean of the other variables involved in the model for the indicator concerned. The partial linear term is subsequently transformed into a probability, rate, or mean using the inverse link functions described in Section 2. The resulting class-specific partial means (probabilities) show the x-y effects on a more natural scale (probability, rate, or mean) for a person with average values on all other variables appearing in the model for the response variable concerned.[41] Note that in models with a single latent variable, without effects of covariates on y_{it}, and without local dependencies involving y_{it}, the resulting partial means equal the means appearing in the model densities, as well as the marginal means. This also applies to the joint partial means in the DFactor model.

Standard errors for the marginal latent probabilities, and the class-specific marginal and partial means are computed using the delta method (see equation 23).

For zero-truncated and zero-censored response variables, the reported class-specific marginal means are expected values conditional on truncation or censoring at $y_{it} = 0$ (see also subsection 8.1.5). In contrast, the reported partial means are expected values unconditional on truncation or censoring.

[41] For nominal variables, we fill in the unweighted average category-specific effects.

As mentioned above, Latent GOLD also provides Profile output for covariates. It reports the class-specific means for numeric covariates, as well as the class-specific probabilities of being in a certain range of covariate values. These means and probabilities are obtained by aggregating and re-scaling posterior membership probabilities:

$$\begin{aligned}\widehat{P}(z_{ir}=a|x) &= \frac{\sum_{z_{ir}=a} w_i\,\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i)}{\sum_i w_i\,\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i)}\\ \widehat{E}(z_{ir}|x) &= \frac{\sum_{i=1}^{I} z_{ir}\,w_i\,\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i)}{\sum_i w_i\,\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i)}\end{aligned}$$

It should be noted that Profile output for covariates is, in fact, re-scaled ProbMeans output (for more details, see ProbMeans output below).

The (joint) partial and marginal conditional means are plotted in the Profile Plot. This makes it easy to identify differences between Clusters, Classes, or levels of DFactors.

8.4 ProbMeans

The Profile output gives the conditional distribution of a y variable given an individual's score on a latent variable. However, it may also be of interest to look at the distribution of the latent variable for a certain level of an observed variable. Measures of these kinds appear in the ProbMeans output and the associated Uni-, Bi-, and Tri-Plots.

The probability of being classified in a certain latent class given a particular z or y value – $\widehat{P}(x|z_{ir}=a)$ or $\widehat{P}(x|y_{it}=a)$ – can be obtained by aggregating the latent classification probabilities defined in equation (25) in the appropriate manner. More precisely,

$$\widehat{P}(x|y_{it}=a)=\frac{\sum_{y_{it}=a} w_i\,\widehat{P}(x|\mathbf{z}_i,\mathbf{y}_i)}{\sum_{y_{it}=a} w_i}.$$

A similar formula can be used to obtain the probabilities for the covariates, $\widehat{P}(x|z_{ir}=a)$.

In LC Cluster and Regression models, the above average class membership probabilities are plotted in Uni- and Tri-plots (Magidson and Vermunt, 2001; Vermunt and Magidson 2000). Similar plots have been proposed by Van der Ark and Van der Heijden (1998) and Van der Heijden, Gilula, and Van der Ark (1999) for standard LC and latent budget models.

In LC DFactor models, we work with DFactor means rather than class membership probabilities; that is,

$$\widehat{E}(x_\ell|y_{it}=a)=\sum_{x_\ell=1}^{K_\ell} x_{x_\ell}^{\ell*}\,\widehat{P}(x_\ell|y_{it}=a).$$

Here, $x_{x_\ell}^{\ell*}$ is the category score of level x_ℓ of the ℓth DFactor. The same formula can be used to obtain a DFactor mean for a certain covariate value. Note that in LC DFactor analysis we use scores ranging from 0 to 1 for the DFactor levels. This implies that with a dichotomous DFactor, a DFactor mean equals the probability of being in level (class) 2.

The DFactor means for each covariate level and for each indicator level are plotted in the Uni- and Bi-plots of our DFactor models (Magidson and Vermunt, 2001; Vermunt and Magidson 2000).

A nice feature of the ProbMeans output is that it describes the relationships between the latent variable(s) and all variables selected as indicators or covariates. This means that even if a certain covariate effect is fixed to zero, one still obtains its ProbMeans information. This feature is exploited in the "inactive covariates method". An advantage of working with inactive covariates is that they do not influence the obtained solution.

8.5 Frequencies / Residuals

In LC Cluster and DFactor models containing only discrete (nominal, ordinal, or Poisson/binomial count) indicators and in LC Regression models for a nominal, ordinal, or (Poisson/binomial) count dependent variable, Latent GOLD reports estimated and observed cell frequencies ($\widehat{m}_{i^*}$ and n_{i^*}), as well as standardized residuals ($\widehat{r}_{i^*}$). Note that this information is provided only for the observed data patterns (the non-zero observed cells).[42] The computation of the estimated cell entries was described in equation (24). The standardized residuals are defined as

$$\widehat{r}_{i^*}=\frac{\widehat{m}_{i^*}-n_{i^*}}{\sqrt{\widehat{m}_{i^*}}}.$$

Note that $(\widehat{r}_{i^*})^2$ is cell i^*'s the contribution to the X^2 statistic.

This output section also contains a column *Cook's D (Cook's Distance)*. This measure can be used to detect influential cases or, more precisely, cases

[42]The option WriteExemplaryData in Latent GOLD Syntax yields the estimated frequencies for all possible data patterns.

having a larger influence on the parameter estimates than others. The exact formula that is used in Latent GOLD 5.1 is given in equation (31). A typical cut-point for Cook's D is four times the number of parameters divided by the number of cases (Skrondal and Rabe-Hesketh, 2004). Note that the reported value in a particular row corresponds to the Cook's D for each of the cases with that data pattern.

8.6 Bivariate Residuals

As explained when presenting the various types of LC models implemented in Latent GOLD, one of the main assumptions in LC analysis is local independence. The reported bivariate residuals for z^{cov}-y and y-y pairs provide direct checks of this assumption. They indicate how similar the estimated and observed bivariate associations are. The measures can be interpreted as lower bound estimates for the improvement in fit (L^2 or $-2\ \log \mathcal{L}$) if the corresponding local independence constraints were relaxed.[43]

Actually, these measures, which are sometimes referred to as modification indices, are Score- or Lagrange-type chi-squared statistics. *Likelihood-ratio* statistics compare the log-likelihood values of a restricted and an unrestricted model, *Wald* statistics estimate the decrease of the log-likelihood value if constraints are imposed in the unrestricted model, and *Score* or *Lagrange-multiplier* statistics estimate the increase of the log-likelihood value if constraints are relaxed in the restricted model or, equivalently, if parameters are added to the restricted model (Buse, 1982). Even though the bivariate residuals reported in Latent GOLD are similar to Score or Lagrange-multiplier tests, they are not exactly the same because they do not take the dependencies between the parameters in the restricted model and the new set into account (Oberski, Van Kollenburg, Vermunt, 2013).

Suppose that a particular local dependency contains P parameters denoted by ϑ_p^{local}. The general formula for our Lagrange-type residuals is:

$$BVR = \frac{1}{P}\sum_{p=1}^{P}\left(\frac{\partial \log \mathcal{P}}{\partial \vartheta_p^{local}}\right)^2 \Big/ \left(\frac{\partial^2 \log \mathcal{P}}{\partial^2 \vartheta_p^{local}}\right).$$

Thus, for each of the P parameters, the first- and second-order derivatives

[43] Similar but not completely equivalent approaches for detecting local dependencies have been proposed by Bartholomew and Tzamourani (1999), Glas (1999), Reiser (1996), and Reiser and Lin (1999)

of the log-posterior is computed. Dependencies between the parameters in a set, as well as with the main effects, is dealt with by including redundant parameters in a set. To further reduce the computational burden, in some situations, we use an approximation for the second-order derivatives. Note that because of the factor $\frac{1}{P}$, the measure should be interpreted as the estimated model fit improvement per extra parameter.

Latent GOLD reports z^{cov}-y residuals for all types of y variables. These are based on the relevant set of β_r^t (β_{mr}^t with a nominal y) terms. The y-y residuals for pairs of categorical (nominal or ordinal) variables are based on $\beta_{mm'}^{tt'}$ (nominal-nominal), $\beta_{.m'}^{tt'}$ (ordinal-nominal), $\beta_{m.}^{tt'}$ (nominal-ordinal), or $\beta_{..}^{tt'}$ (ordinal-ordinal) terms, and for pairs of continuous variables on the error covariances $\sigma_{tt'}$. These are all terms that can be included in the model.

For other y-y pairs, we use a more ad hoc procedure to determine the size of the residuals. Residuals for categorical-continuous and categorical-count pairs assume that the categorical variable is (also) used as a covariate affecting the continuous or count variable. Count-continuous residuals use the count as a covariate in the model for the continuous indicator, and count-count residuals are obtained as the average BVR of the regression of y_{it} on $y_{it'}$ and $y_{it'}$ on y_{it}. As explained in the subsection on LC Cluster models for mixed mode data, these types of local dependencies can only be included in the model by a trick that requires including a variable both as covariate and as indicator in the model.

For categorical indicators, a BVR has the form of a Pearson chi-squared divided by P; that is,

$$BVR_{t,t'} = \frac{1}{P} \sum_{m=1}^{M_t} \sum_{m'=1}^{M_{t'}} \frac{[n_{m,m'} - E(n_{m,m'})]^2}{E(n_{m,m'})},$$

where

$$E(n_{m,m'}) = \sum_{i=1}^{I} w_i \sum_{x=1}^{K} \widehat{P}(y_{it} = m|x) \widehat{P}(y_{it'} = m'|x) \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i).$$

Note that even though not indicated explicitly, the sum i is over the cases for which both y_{it} and $y_{it'}$ are observed. Also the observed frequency table with entries $n_{m,m'}$ is obtained with the cases without missing values on y_{it} and $y_{it'}$. For ordinal indicators, we replace $n_{m,m'}$ by $\widehat{n}_{m,m'}$. These are frequency counts respecting the linear-by-linear or linear-by-nominal association structure, obtained with a set of uni-dimensional Newton iterations.

For a covariate and a categorical indicator, the BVR has the same form, with

$$E(n_{m,m'}) = \sum_{i=1}^{I} w_i \sum_{x=1}^{K} \widehat{P}(y_{it} = m|x) I(z_{ip} = m') \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i).$$

Here, $I(z_{ip} = m')$ is an indicator variable taking on the value 1 if the condition is true and otherwise 0. As for ordinal indicators, the observed frequency counts are adjusted when the covariate is numeric.

If the *Bootstrap Chi*^{2} option is used, the program also provides the estimated bootstrap p-values for the BVRs, which are statistics for which the asymptotic distribution is unknown (Oberski, Van Kollenburg, and Vermunt 2013). Also the 5% critical values (CV) and the Monte Carlo standard errors of the p-values are reported.

8.7 Estimated Values

In Regression the Estimated Values output section reports the estimated Class-specific, the estimated overall, and the observed means (probabilities) for each unique predictor pattern. Let $\mathbf{z}_u^{pred}$ refer to a unique predictor pattern. With continuous variables one gets the estimated means $\widehat{\mu}_{t,x,\mathbf{z}_u^{pred}}$; with Poisson counts the estimated rates $\widehat{\theta}_{t,x,\mathbf{z}_u^{pred}}$, with binomial counts the success probabilities $\widehat{\pi}_{t,x,\mathbf{z}_u^{pred}}$, with ordinal or nominal responses the probabilities $\widehat{\pi}_{m|t,x,\mathbf{z}_u^{pred}}$, and with ordinal variables also the means $\sum_{m=1}^{M_t} y_m^{t*} \widehat{\pi}_{m|t,x,\mathbf{z}_u^{pred}}$. The overall estimates are weighted averages of the Class-specific estimates.

For zero-truncated and zero-censored response variables, the class-specific estimated values are conditional on truncation or censoring at $y_{it} = 0$ (see also subsection 8.1.5).

In Cluster, DFactor and Step3, Estimated Values contains the model probabilities and means. In most models this will be the same information as provided by Classification-Model and Profile; that is, the class membership probabilities conditional on covariates and the class-specific indicator probabilities or means. However, Estimated Values gives more detailed information in DFactor models with multiple DFactors, in models with direct effects of covariates on indicators, and in models with direct effects between categorical indicators. These are situations in which Classification-Model and Profile give marginal probabilities and means.

8.8 Classification

The *Classification-Posterior* output section contains the classification information for each response pattern i^*. Classification is based on the latent classification or posterior class membership probabilities, $\widehat{P}(x|\mathbf{z}_{i^*}, \mathbf{y}_{i^*})$, described in equation (25). In LC Cluster models and LC Regression models, these quantities are used to determine to which latent class someone belongs. More precisely, subjects are assigned to the latent class with the highest latent classification probability. This method of assignment is sometimes referred to a empirical bayes modal (EBM) or modal a posteriori (MAP) estimation (Skrondal and Rabe-Hesketh, 2004).

In DFactor models, there may be more that one latent variable. In addition, the discrete factors are ordinal (or discrete interval) variables that can be used as approximations of continuous latent variables with unknown distributions. Therefore, in LC DFactor models, Latent GOLD not only reports the classification probabilities $\widehat{P}(x_\ell|\mathbf{z}_{i^*}, \mathbf{y}_{i^*})$ and the modal allocation for each DFactor, but also the estimated DFactor scores or posterior DFactor means. The DFactor scores for response pattern i are obtained by

$$\widehat{E}(x_\ell|\mathbf{z}_{i^*}, \mathbf{y}_{i^*}) = \sum_{x_\ell=1}^{K_\ell} x_{x_\ell}^{\ell*} \widehat{P}(x_\ell|\mathbf{z}_{i^*}, \mathbf{y}_{i^*}).$$

Here, $x_{x_\ell}^{\ell*}$ denote the category scores (ranging from 0 to 1) for the levels of the latent variable concerned. DFactor scores computed in this manner are sometimes referred to as empirical bayes (EB) or expected a posterior (EAP) estimators (Skrondal and Rabe-Hesketh, 2004).

Classification-Model is the classification that is obtained based on covariates only. This involves using the model probabilities $\widehat{P}(x|\mathbf{z}_u)$, sometimes referred to as prior probabilities, as classification probabilities for each covariate pattern u. In DFactor models, we need to compute the DFactor-specific marginal conditional probabilities $\widehat{P}(x_\ell|\mathbf{z}_u)$. The classification rules are the same as applied to the posterior class membership probabilities.

8.9 Output-to-file Options

Five different types of information can be output to a data files: classification, classification based on covariates, predicted values, individual-specific coefficients, and the estimated variance-covariance matrix of the model parameters.

For *Classification-Posterior* and *Classification-Model*, the output file contains the posterior class-membership probabilities $\widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)$ and the model probabilities $\widehat{P}(x|\mathbf{z}_i)$, respectively, as well as the modal class assignments based on these probabilities. For DFactor models, the file contains the DFactor-specific classification probabilities, the DFactor means, and the joint classification probabilities for all DFactors simultaneously.

For Regression models, the *Predicted Values* can be output to a file. The file contains the estimated individual-specific predicted values $\widehat{y}_{it}$ and/or individual-specific probabilities $\widehat{P}_{m|it}$, which were defined in equations (28) and (29). Instead of posterior mean predictions, one can also request HB-like predictions or model-based predictions.

For *Individual Coefficients*, the file contains the estimated individual-specific regression coefficients. Let $\widehat{\beta}_{xq}$ denote the estimated value of one of the regression coefficients (intercept or predictor effect). The posterior-mean or expected a posteriori estimate of regression coefficient q for case i is defined as follows:

$$\widehat{\beta}_{iq} = \sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\, \widehat{\beta}_{xq} \tag{30}$$

that is, as a weighted average of the Class-specific coefficients. These estimates are similar to the individual coefficients obtained in multilevel, mixed-effects, random-effects, and hierarchical Bayes (HB) models. The person-specific coefficients can be used to predict person i's new responses. The posterior standard deviations are defined as

$$\widehat{\sigma}_{\widehat{\beta}_{iq}} = \sqrt{\sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i)\, \left(\widehat{\beta}_{xq} - \widehat{\beta}_{iq}\right)^2}$$

Another output-to-file item is *Cook's D (Cook's Distance)*. It can be used to detect influential cases or, more precisely, cases with a large influence on the parameter estimates. The formula that is used the following:

$$C_i = -2\, \mathbf{g}_i'\, \mathbf{H}^{-1}\, \mathbf{g}_i, \tag{31}$$

where $\mathbf{H}$ is the Hessian matrix and $\mathbf{g}_i$ the vector with the gradient contributions of case i. A typical cut-point for Cook's D is four times the number of parameters divided by the number of cases (Skrondal and Rabe-Hesketh, 2004).

In Step3-Covariate and Step3-Scoring one can write the scoring equations (the logistic equation for the class memberships) to an SPSS syntax file or to a more generic syntax that can easily be adapted for other packages.

The last output-to-file item is the *Variance-Covariance Matrix* of the model parameters. Dependent of the type of variance estimator that is requested this will be $\widehat{\boldsymbol{\Sigma}}_{standard}(\boldsymbol{\vartheta})$, $\widehat{\boldsymbol{\Sigma}}_{outer}(\boldsymbol{\vartheta})$, or $\widehat{\boldsymbol{\Sigma}}_{robust}(\boldsymbol{\vartheta})$. Note that also the variances and covariances involving the omitted categories of the effect coded nominal variables are reported.

Part II: Advanced Model Options, Technical Settings, and Output Sections

9 Introduction to Part II: Advanced Models

This part of the manual describes the four Advanced options of Latent GOLD 5.1. These are:

1. The Markov submodule for the estimation of the most important types of latent Markov models,

2. An option for specifying models with continuous latent variables, which are referred to as continuous factors (CFactors).

3. A multilevel extension of the LC/FM model, which is a model containing group-level continuous latent variables (GCFactors) and/or a group-level nominal latent variable (GClasses).

4. An option to deal with the sampling design, which yields correct statistical tests for complex survey sampling designs that deviate from simple random sampling.

The submodule called *Markov* can be used for the estimation of many types of latent Markov models. These are LC models for longitudinal data in which individuals are allowed to switch between latent classes across measurement occasions. Latent Markov models are also referred to as latent transition models, hidden Markov models, Markov-switching models, or regime-switching models. The implementation in Latent GOLD Advanced allows inclusion of time-constant and time-varying covariates, the specification of mixture variants, including mover-stayer models, and the use of multiple indicators of different scale types. By using an efficient EM algorithm, Latent GOLD's Markov submodule can deal with models for very large numbers of time points.

The *Continuous Factors* (CFactors) option can be used to specify factor-analytic models, item response theory (IRT) models, and random-effects regression models for two-level data, where the same options concerning the

scale types of the indicators and the dependent variable apply as in other Latent GOLD models. One may also combine CFactors with a single nominal latent variable (Cluster and Regression submodules) or with multiple ordinal latent variables (DFactor submodule), yielding too many interesting special cases to be listed here. One example is a standard LC model for categorical indicators in which local dependencies are captured by a continuous latent variable. Another example is a LC or FM Regression model, in which the intercept is allowed to vary within latent classes by including a random intercept in the model. In an IRT or factor-analytic model, one could introduce latent classes to capture unobserved population heterogeneity in some of the model parameters.

The *Multilevel Model* option can be used to define LC and FM models for nested data, such as employees nested within firms, pupils nested within schools, clients nested within stores, patients nested within hospitals, citizens nested within regions, and repeated measurements nested within individuals. Note that LC and FM models are themselves models for two-level data; that is, models for multiple responses per case. The multilevel LC (FM) model is thus, in fact, a model for three-level data; that is, for multiple responses nested within cases and cases nested within groups. As in any multilevel analysis, the basic idea of a multilevel LC analysis is that one or more parameters of the model of interest is allowed to vary across groups using a random-effects modeling approach. In Latent GOLD, the group-level random effects can either be specified to be continuous (group-level continuous factors: GCFactors) or discrete (group-level latent classes: GClasses), yielding either a parametric or a nonparametric approach, respectively.

One variant of the multilevel LC model involves including group-level random effects in the model for the latent classes, which is a way to take into account that groups differ with respect to the distribution of their members across latent classes (Vermunt, 2003, 2005; Vermunt and Magidson, 2005b). Not only the intercept, but also the covariate effects may have a random part. In Cluster and DFactor models, it is also possible to allow GCFactors and/or GClasses to have direct effects on the indicators. In Regression, one can include GCFactors and GClasses in the model for the dependent variable. By combining group-level with case-level latent classes, one obtains a three-level generalized linear model with nonparametric random effects, and by combining group-level continuous factors with case-level continuous factors one obtains a standard three-level random-coefficients generalized linear model (Vermunt, 2002c, 2004).

The *Survey* option makes it possible to get correct statistical tests with stratified and clustered samples, as well as with sampling weights and samples from finite populations. This option applies to any model that can be estimated with Latent GOLD. The design-corrected variance-covariance matrix of the model parameters is obtained by the well-known linearization estimator. Sampling weights can also be dealt with using a two-step procedure that involves estimating the model without sampling weights, and subsequently correcting the latent class distribution and covariate effects using the sampling weights.

The next four sections describe the four Advanced options in more detail. Attention is paid to model components, estimation issues, and application types. The last section discusses the output obtained with the Latent GOLD Advanced options.

10 Latent Markov Models

The Markov submodule which is part of Latent GOLD 5.1 Advanced can be used to estimate latent Markov models, including mixture variants and models with covariates (Bartolucci, et al. 2007, 2013); Kaplan, 2008; Paas, Vermunt and Bijmolt, 2007; Poulsen, 1982, 1990; Van de Pol and De Leeuw, 1990; Vermunt, Langeheine, and Böckenholt, 1999; Vermunt, Tran, and Magidson, 2008; Vermunt, 2010;. A latent Markov model is a LC model for longitudinal data in which persons are allowed to switch between latent classes across measurement occasions. It is also referred to as latent transition model (Collins and Lanza, 2010; Collins and Wugalter, 1992), hidden Markov model (MacDonald and Zucchini, 1997; Visser, 2011), Markov-switching (Frühwirth-Schnatter, 2009), or regime-switching model (Hamilton, 2008).

Below, we first introduce some notation and describe the simplest latent Markov model. Then, we describe the general model implemented in Latent GOLD Advanced, including the use of transition coding for the logit parameters of the transition probability model. Subsequently, we present various constrained models and provide the relevant details on parameter estimation using the Baum-Welch algorithm.

10.1 The Simplest Latent Markov Model

To distinguish the dynamic class membership from the static class membership in standard mixture models, we refer to the dynamic classes as latent states. Let x_t^d denote the latent state at time point t, where superscript d refers to dynamic and is used to distinguish this variable from the other types of latent variables (case- and group-level classes). The time variable t runs from 0 to T_i, the last measurement occasions for the person concerned.[44]

Let us first look at a latent Markov model for a single categorical response variable y_{it}, where the vector collecting the responses of person i at all time points is denoted by $\mathbf{y}_i$. This simple latent Markov model is defined as follows:

$$P(\mathbf{y}_i) = \sum_{x_0^d=1}^{K^d} \sum_{x_1^d=1}^{K^d} ... \sum_{x_T^d=1}^{K^d} P(x_0^d) \prod_{t=1}^{T_i} P(x_t^d|x_{t-1}^d) \prod_{t=0}^{T_i} P(y_{it}|x_t^d).$$

The three sets of probabilities defining this model are the initial state probabilities $P(x_0^d)$, the transition probabilities $P(x_t^d|x_{t-1}^d)$, and the response probabilities $P(y_{it}|x_t^w)$.

The latent Markov model is a constrained LC model with $T_i + 1$ latent variables. The two main assumptions of the latent Markov model can be derived from the above equation. The product $\prod_{t=1}^{T_i} P(x_t^d|x_{t-1}^d)$ follows from the first-order Markov assumption, indicating that the latent state at time point t depends on the state at $t-1$, but not on the states at earlier time points. The product $\prod_{t=0}^{T_i} P(y_{it}|x_t^d)$ indicates that the response at time point t depends on the latent state at this time point, but not on the latent states or the responses at other time points. The latter assumption is sometimes referred to as assumption of independent classification errors (ICE).

10.2 The General Mixture Latent Markov with Covariates

The general model implemented in Latent GOLD Advanced, expands the simple latent Markov model in various important ways. It allows inclusion

[44]Note that in other models, we use t to denote a particular indicator or replication and T and T_i to denote the number of indicators and the number of replications, implying that our notation is somewhat inconsistent. However, we prefer using the common notation t and T_i to refer to the (number of) time points. This means that in Markov other symbols are needed to denote the (number of) indicators.

of a mixture variable x (Poulsen, 1990; Van de Pol and Langeheine, 1990), inclusion of covariates $\mathbf{z}_i$ (Vermunt, Langeheine, an Böckenholt, 1999), and having multiple indicators of any of the Latent GOLD scale types (see, for example, Schmittman et al., 2005). This general model has the following form

$$\begin{aligned} f(\mathbf{y}_i|\mathbf{z}_i) &= \sum_{x=1}^{K} \sum_{x_0^d=1}^{K^d} \sum_{x_1^d=1}^{K^d} ... \sum_{x_T^d=1}^{K^d} P(x|\mathbf{z}_i) P(x_0^d|x, \mathbf{z}_i) \prod_{t=1}^{T_i} P(x_t^d|x_{t-1}^d, x, \mathbf{z}_{it}) \\ & \prod_{t=0}^{T_i} f(\mathbf{y}_{it}|x_t^d, x, \mathbf{z}_{it}). \end{aligned} \tag{32}$$

As can be seen, here we have four sets of probabilities: $P(x|\mathbf{z}_i)$ are class proportions which may depend on time-constant covariates, $P(x_0^d|x, \mathbf{z}_i)$ are initial state probabilities which may depend on classes and time-constant covariates, $P(x_t^d|x_{t-1}^d, x, \mathbf{z}_{it})$ are transition probabilities which may depend on classes and time-varying covariates, and $f(\mathbf{y}_{it}|x_t^d, x, \mathbf{z}_{it})$ are indicator distributions which may depend on latent states, classes, and time-varying covariates.

The distribution of the indicators is modeled in the same way as in Cluster models, where it should be noted that in the default specification indicators are assumed to be independent of x and $\mathbf{z}_{it}$ and mutually independent given x_t^d. The class, initial state, and transition probabilities are modeled using logistic regression equations. For the classes and initial states, these are a standard logistic regression models; for the transitions, this can be either a standard logistic regression model or a model we refer to as a transition logit model (see option on the Advanced tab). We will not discuss the standard logistic regression models here, but instead focus on the less common but very useful transition logit model, which has the following form:

$$\log \frac{P(x_t^d = s|x_{t-1}^d = r, x, \mathbf{z}_{it})}{P(x_t^d = r|x_{t-1}^d = r, x, \mathbf{z}_{it})} = \gamma_{xrs0} + \sum_{p=1}^{P} \gamma_{xrsp} \cdot z_{itp},$$

for $r \neq s$. As can be seen, we model the logit comparing the probability of moving from origin state r to destination state s with the probability of staying in state r, which is what we refer to as a transition logit. The model has a separate γ parameter for each transition logit, whereas the parameters for $r = s$ (no transition) are fixed to 0 for identification. Note that this implies that we use dummy coding for the "response" variable x_t^d, but with

a reference category that depends on (changes with) the value of the origin state x^d_{t-1}. We call this type of coding "transition coding".

In mixture latent Markov models, which are defined with the Advanced options in the Markov submodule, the γ parameters can be allowed to depend on the classes x, which is indicated with the index x. Also in the regression parameters for the initial state probabilities can be allowed to vary across classes.

10.3 Restrictions

Four types of restricted (mixture) latent Markov models can be defined using the Advanced options. One important restriction involves defining the first class to be a *stayer class*, that is, a class with transition probabilities restricted to 0. More formally, $P(x^d_t = s|x^d_{t-1} = r, x = 1, \mathbf{z}_{it}) = 0$ for $r \neq s$. This yields what is often referred to as a mover-stayer model. For an application, see Magidson, Vermunt, and Tran (2009).

The second restriction involves defining the states to be perfectly related to one of the categorical indicators (the first one for which $K^d = M$); that is, $P(y_{it} = m|x^d_t = s) = 0$ for $m \neq s$. This is a trick to make it possible to use the Markov submodule for the estimation of simple *manifest Markov models* or *mixture Markov models*. For an application of the latter, see Dias and Vermunt (2007).

The last two restrictions concern the scale type of the latent state variable. The ordinal option defines a model with latent states with fixed equidistant scores which are used to restrict the state-indicator associations, which is similar to a one-DFactor model. The option order-restricted defines a model imposing monotonicity restrictions on the state-indicator associations, a restriction that is also available in the Cluster submodule (where it is called order-restricted clusters).

10.4 Parameter Estimation

ML (PM) estimates are found by a combination of EM and Newton Raphson iterations. In order to be able to deal with applications involving large numbers of time points (see, e.g., Ramos, Vermunt, and Dias, 2011), the E step computations use a generalized version of the forward-backward recursion scheme, also known as the Baum-Welch algorithm, originally proposed by Baum et al. (1970). Details on this generalized Baum-Welch algorithm,

which can deal with mixtures, covariates, and multiple response variables, are provided by Vermunt, Tran, Magidson (2008). For the Newton-Raphson iterations, analytic first- and second-order derivatives are computed using the forward recursion scheme described by Lystig and Hughes (2002).

It may be useful for users to know that in developing the Markov submodule, we decided to use the existing Latent GOLD Syntax system to estimate the Markov models; that is, based on the model specified via point-and-click GUI, we generate and run the corresponding LG-Syntax model. Therefore, the output formatting is very similar to that of the Syntax models. Two differences are that the restricted parameters for the Stayer class and the Perfect states are not reported in Parameters and that Classes and States are reordered (from large to small) so that they appear in the same order when rerunning a model.

11 Continuous Factors

11.1 Model Components and Estimation Issues

Let F_{di} denote the score of case i on continuous latent variable, factor, or random effect d. The total number of CFactors is denoted by D, $1 \leq d \leq D$, and the full vector of CFactor scores by $\mathbf{F}_i$. The maximum number of CFactors that can be included in a Latent GOLD model is three, thus $0 \leq D \leq 3$.

Recall that without CFactors the most general Latent GOLD structure for $f(\mathbf{y}_i|\mathbf{z}_i)$ equals

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{x=1}^{K} P(x|\mathbf{z}_i)\, f(\mathbf{y}_i|x, \mathbf{z}_i),$$

where

$$f(\mathbf{y}_i|x, \mathbf{z}_i), = \prod_{h=1}^{H} f(\mathbf{y}_{ih}|x, \mathbf{z}_i).$$

If we include CFactors in a model, the assumed structure for $f(\mathbf{y}_i|\mathbf{z}_i)$ becomes

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{x=1}^{K} \int_{\mathbf{F}_i} f(\mathbf{F}_i)\, P(x|\mathbf{z}_i)\, f(\mathbf{y}_i|x, \mathbf{z}_i, \mathbf{F}_i)\, \mathrm{d}\mathbf{F}_i, \tag{33}$$

where

$$f(\mathbf{y}_i|x,\mathbf{z}_i,\mathbf{F}_i) = \prod_{h=1}^{H} f(\ \mathbf{y}_{ih}|x,\mathbf{z}_i,\mathbf{F}_i)$$

The F_{di} are assumed to be standard normally distributed and mutually independent. In other words, $f(\mathbf{F}_i) = N(\mathbf{0},\mathbf{I})$, where $\mathbf{I}$ is the identity matrix. As will be shown below, this specification is much less restrictive than one may initially think.

It is also possible to define models – e.g., IRT models or random-effects regression models – containing CFactors but no latent classes x. That simplifies the structure for $f(\mathbf{y}_i|\mathbf{z}_i)$ to

$$f(\mathbf{y}_i|\mathbf{z}_i) = \int_{\mathbf{F}_i} f(\mathbf{F}_i)\ f(\mathbf{y}_i|\mathbf{z}_i,\mathbf{F}_i)\,\mathrm{d}\mathbf{F}_i,$$

with

$$f(\mathbf{y}_i|\mathbf{z}_i,\mathbf{F}_i) = \prod_{h=1}^{H} f(\ \mathbf{y}_{ih}|\mathbf{z}_i,\mathbf{F}_i)$$

Equation (33) shows that the F_{di} may appear in the model for the response variables, but not in the model for the latent classes.[45] In Cluster models, this is accomplished by expanding the linear predictor with the term $\sum_{d=1}^{D}\lambda_{xd}^{t}\cdot F_{di}$, where λ_{xd}^{t} denotes the effect of F_{di} on indicator y_{it} in latent class x. Note that for nominal indicators, λ_{xd}^{t} is replaced by λ_{mxd}^{t}. As can be seen, each CFactor may be related to each indicator, and these relationships may differ across latent classes. In the DFactor submodule, the term corresponding to the CFactors is the same as in Cluster, except that the effects can not be class dependent.

In the Regression submodule, the linear predictor is expanded with the term $\sum_{d=1}^{D}\lambda_{x0d}\cdot F_{di} + \sum_{d=1}^{D}\sum_{q=1}^{Q}\lambda_{xqd}\cdot F_{di}\cdot z_{itq}^{pred}$. For nominal dependent variables, λ_{xqd} is replaced by λ_{xmqd} $(0 \le q \le Q)$. As can be seen, F_{di} may have a direct effect on the dependent variable, which makes it possible to define random-intercept models. The $F_{di}\cdot z_{itq}^{pred}$ product term defines a random coefficient for the predictor concerned. An important difference with the more standard specification of random-effects models is that here each F_{di} can serve as random effect for each of the model effects, which, as will be shown below, can be used to define parsimonious random-effects covariance structures. Another important difference is, of course, that the size of

[45] There is a trick for including CFactor effects in the model for the latent classes using the multilevel option.

the parameters associated with the random effects may differ across latent classes.

Model restrictions In both Cluster and DFactor, it is possible to equate the λ's across indicators of the same type for selected CFactors ("Equal Effects"), as well as to fix some of the λ's to zero. Moreover, in Cluster, the λ's may be assumed to be equal across Clusters ("Cluster Independent").

In Regression, one can use the parameter constraints "Class Independent", "No Effect", and "Merge Effects", implying equal λ's among all Classes, zero λ's in selected Classes, and equal λ's in selected Classes.

ML (PM) estimation and technical settings The main complication in the ML (PM) estimation of models with CFactors is that we have to deal with the multidimensional integral appearing in the definition of the marginal density $f(\mathbf{y}_i|\mathbf{z}_i)$ (see equation 33). A closed form expression for this integral is only available when all response variables are continuous (and normally distributed). In all other situations, the integrals must be solved using approximation methods.

In Cluster and DFactor models containing only continuous indicators, we make use of the fact that the indicators remain multivariate normally distributed even in models with CFactors, but with a variance-covariance matrix equal to $\mathbf{\Lambda}_x\mathbf{\Lambda}_x' + \mathbf{\Sigma}_x$, where $\mathbf{\Lambda}_x$ is a T-by-D matrix containing the factor loadings λ_{xd}^t. In fact, the marginal density $f(\mathbf{y}_i|\mathbf{z}_i)$ has the form of a standard LC or FM model; that is,

$$f(\mathbf{y}_i|\mathbf{z}_i) = \sum_{x=1}^{K} P(x|\mathbf{z}_i)\, f(\mathbf{y}_i|x, \mathbf{z}_i),$$

with $f(\mathbf{y}_i|x, \mathbf{z}_i) = N(\boldsymbol{\mu}_{x,\mathbf{z}_i}, \mathbf{\Lambda}_x\mathbf{\Lambda}_x' + \mathbf{\Sigma}_x)$.

In all other cases, Latent GOLD approximates the conditional density $f(\mathbf{y}_i|\mathbf{z}_i)$ by means of Gauss-Hermite numerical integration, implying that the multidimensional integral is replaced by multiple sums (Bock and Aitkin, 1981). With three CFactors and B quadrature nodes per dimension, the approximate density equals

$$f(\mathbf{y}_i|\mathbf{z}_i) \approx \sum_{x=1}^{K} \sum_{b_1=1}^{B} \sum_{b_2=1}^{B} \sum_{b_3=1}^{B} P(x|\mathbf{z}_i)\, f(\mathbf{y}_i|x, \mathbf{z}_i, F_{b_1}, F_{b_2}, F_{b_3})\, P_{b_1}\, P_{b_2}\, P_{b_3}.$$

Here, F_{b_d} is the location and P_{b_d} the weight corresponding to quadrature node b_d for CFactor d. These nodes and weights are obtained from published quadrature tables (Stroud and Secrest, 1966). As can be seen, because of the multiple sums, this approximate density is very similar to the density of a LC model with multiple latent variables. The above approximation also shows that – given the fact that one will usually use at least 10 quadrature points per dimension (Lessafre and Spiessens, 2001) – because of computation burden, it does not make sense to have models with more than three CFactors.

Similarly to what Latent GOLD does for standard LC and FM models, the ML (PM) estimation problem for models with CFactors is solved using a combination of EM and Newton-Raphson with analytic first- and second-order derivatives.

The only additional technical setting for models with CFactors is the option for specifying the number of quadrature nodes to be used in the numerical integration. The default value is 10, the minimum 2, and the maximum 50.

11.2 Application Types

11.2.1 Factor analysis

One of the applications of the CFactors option is ML factor analysis with continuous indicators. The factor-analytic model implemented in Latent GOLD equals

$$\mu_{t,\mathbf{F}_i} = \beta_0^t + \sum_{d=1}^{D} \lambda_d^t \cdot F_{di}.$$

Special features are that there may be missing values on the indicators, local dependencies (correlated errors) among indicators can be included, loadings may be equated across indicators for selected CFactors ($\lambda_d^t = \lambda_d$), and Heywood cases are prevented from occurring by nonnegativity constraints on the error variances ($\sigma_t^2 \geq 1.0\text{e}{-6} \cdot s_t^2$, were s_t^2 is the sample variance of indicator t).

Two limitations of the Latent GOLD factor-analytic model are that CFactors are assumed to be uncorrelated, as well as that there are no rotation options. The latter means that the solutions must be identified by fixing one or more factor loadings to zero and/or by equating loadings across indicators

for selected factors.[46]

Whereas ML factor analysis can be performed with any structural equation modeling (SEM) package, the unique feature of Latent GOLD is that it can be combined with a LC or FM model. This yields what McLachlan and Peel (2000) refer to as "mixtures of factor analyzers". This model differs from the above factor-analytic model in that its parameter may differ across latent classes or mixture components:

$$\mu_{t,x,\mathbf{F}_i} = \beta_0^t + \beta_{x0}^t + \sum_{d=1}^{D} \lambda_{xd}^t \cdot F_{di}.$$

It should be noted that this is, in fact, a mixture of multivariate normals, in which the within-Cluster covariances between indicators are assumed to be in agreement with a factor-analytic structure. As already explained above, the within-class covariance matrix equals $\mathbf{\Lambda}_x \mathbf{\Lambda}_x' + \mathbf{\Sigma}_x$. This shows that the CFactors option may be used as a tool for defining parsimonious local dependency structures in the context of mixture-model clustering. An even more parsimonious specification is obtained by assuming that the factor loadings are class independent; that is, by setting $\mathbf{\Lambda}_x = \mathbf{\Lambda}$.

In other applications, one may be interested in the factor-analytic model itself, and use the LC or FM model as a tool for studying and detecting unobserved heterogeneity (Yung, 1997).[47] In such cases, one would most probably use a simple confirmatory factor-analytic structure, such as a one-factor model or a two-factor growth model. Assuming that the T indicators are measurements of the same variable at T time points, a LC two-factor growth model is obtained as follows:

$$\mu_{t,x,\mathbf{F}_i} = \beta_0^t + \beta_{x0}^t + \lambda_{x1} \cdot F_{1i} + \lambda_{x2}^t \cdot F_{2i}.$$

where λ_{x1} captures the inter-individual variation in the intercept and λ_{x2}^t the inter-individual variation in the time effect. Note that the loadings corresponding to the first factor are assumed to be equal across indicators ($\lambda_{x1}^t = \lambda_{x1}$).

[46] When $D = 1$, no identifying restrictions need to be imposed. With $D = 2$, one has to impose one constraint, typically fixing one λ_2^t to 0. With $D = 3$, three constraints are needed, typically fixing one λ_2^t and two λ_3^t terms to 0.

[47] This is the most important special case of the more general framework of mixture structural equation modeling (Dolan and Van der Maas, 1997; Jedidi, Jagpal, and DeSarbo, 1997).

11.2.2 IRT models

The main difference between factor analysis and IRT modeling is that in the former response variables are continuous, whereas in the latter they are discrete (Bartholomew and Knott, 1999). With the Latent GOLD CFactors option, it is possible to obtain (marginal) ML estimates for the most important types of parametric IRT models for binary, ordinal, and nominal response variables, as well as for IRT models for mixed responses.[48] Note that if discrete and continuous indicators are mixed, one obtains a hybrid between factor analysis and IRT, which Moustaki and Knott (2000) called a generalized IRT model.

The linear term in a univariate IRT model equals

$$\begin{aligned} \eta^t_{F_i} &= \beta^t_0 + \lambda^t \cdot F_i, \\ \eta^t_{m,F_i} &= \beta^t_{m0} + \lambda^t_{\cdot} \cdot y^{*t}_m \cdot F_i, \\ \eta^t_{m,F_i} &= \beta^t_{m0} + \lambda^t_m \cdot F_i, \end{aligned}$$

for dichotomous, ordinal, and nominal responses, respectively, yielding the two-parameter logistic (Birnbaum, 1968), the generalized partial-credit (Muraki, 1992), and the nominal-response model (Bock, 1972). It should be noted that the parameterization used in IRT modeling is sometimes a bit different from the one used here. For example, the two-parameter logistic model is usually written as

$$\eta^t_{F_i} = a^t\,(F_i - b^t),$$

where b^t and a^t are the difficulty and the discrimination parameter, respectively. These can, however, be obtained easily from the Latent GOLD parameterization: $a^t = \lambda^t$ and $b^t = -\beta^t_0/\lambda^t$. By constraining $\lambda^t = \lambda$ ($a^t = a$), one obtains the one-parameter logistic or Rasch model (Rasch, 1980) in the binary case and the partial-credit model (Masters, 1982) in the ordinal case.[49]

Latent GOLD can deal with a much more general class of IRT models than these standard models. First of all, it is possible to increase the number of dimensions up to three, yielding multidimensional variants of the

[48] Overviews of the field of parametric IRT modeling can be found in Hambleton and Swaminathan (1985), Heinen (1996), Mellenbergh (1995), and Van der Linden and Hamilton (1997).

[49] It is also possible to assume that $a = 1$ and treat the variance of the latent variable as an unknown parameter to be estimated (see also discussion on random-effects models).

IRT models discussed above (Reckase, 1997). Similarly, as explained above in the context of factor analysis, an appropriately "rotated" solution can be obtained by fixing certain discrimination parameters (loadings) to zero. Moreover, contrary to standard IRT models, there is no need to assume that the indicators are of the same scale type. One can not only combine binary, ordinal, and nominal responses, but also use continuous indicators or counts, including their censored and truncated variants. This yields the generalized IRT model proposed by Moustaki and Knott (2000).

The standard IRT models for dichotomous and nominal variables, can not only be specified using the Cluster submodule, but also using the Regression submodule. With T items one defines $T-1$ item dummies – or equivalently, a single nominal predictor variable "item number" – that enter as predictors in a random-effects (logistic) regression model. In fact, the Regression submodule offers a fully general linear-logistic test model that can take into account multiple item-design factors. For ordinal responses, such a logistic regression-based IRT model is more restricted than the one specified with the Cluster submodule because the Latent GOLD ordinal logistic regression procedure will always imply certain restrictions on the intercept across replications. These restrictions turn out to yield a restricted variant of the partial-credit model called the rating-scale model (Andrich, 1978), in which $\beta^t_{m0} = \beta_{m0} + y^{*t}_m \cdot \beta^t_{.0}$.

The unique feature of Latent GOLD is that it allows combining CFactors with latent classes. This yields FM variants of IRT models, such as the mixed-Rasch model proposed by Rost (1990) and the mixed-Birnbaum model proposed by Smit, Kelderman, and Van der Flier (2000). The mixed-Birnbaum model, for example, is defined as

$$\eta^t_{x,F_i} = \beta^t_0 + \beta^t_{x0} + \lambda^t_x \cdot F_i, \tag{34}$$

whereas the mixed-Rasch model is obtained by assuming that $\lambda^t_x = \lambda_x$.

11.2.3 Local dependence LC models

Qu, Tan, and Kutner (1996) and Hadgu and Qu (1998) proposed using IRT-like structures within latent classes in medical applications based on sets of "symptoms" (yes/no responses). In most of the medical applications of LC analysis, the method is used to build a two-class diagnostic instrument distinguishing between normal and abnormal cases. The fact that a two-class model might not fit the data is either seen as an artifact of the measurement

instrument (some symptoms are more similar to one another than to the other ones) or as a result of within-class heterogeneity. It is indeed true that these two quite related requirements (local independence and within-class homogeneity) are strong assumptions that may not hold in a simple two-class model.

To allow for local dependencies and within-class heterogeneity, Qu, Tan, and Kutner (1996) proposed expanding the standard LC model with a CFactor, yielding what they called a random-effects LC model. The model they proposed is a binary logistic model with the linear predictor

$$\eta^t_{x,F_i} = \beta^t_0 + \beta^t_{x0} + \lambda^t \cdot F_i,$$

which is similar to the mixed-Birnbaum model described in equation (34). The difference is that here the λ^t are class independent. More restricted variants were also defined in which the λ^t were equated across all items or across selected sets of items, and in which some λ^t were equated to zero.

11.2.4 Random-effects models

Another important application of CFactors involves random-effects modeling using any of the generalized linear models implemented in the Regression submodule (Agresti, Booth, Hobert, and Caffo, 2000; Goldstein, 1995; Hedeker, 2003; Hedeker and Gibbons, 1996; Snijders and Bosker, 1999; Skrondal and Rabe-Hesketh, 2004; McFadden and Train, 2000; Wong and Mason, 1985).[50]

Let us first look at the *random intercept* model. Denoting the random part of the intercept by Ψ_{0i}, the linear predictor in a random-intercept model for either binary, continuous, or count responses would be of the form:

$$\eta_{\mathbf{z}_{it},\Psi_{0i}} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z^{pred}_{itq} + \Psi_{0i}.$$

The parameters to be estimated are not only the fixed effects β_q $(0 \leq q \leq Q)$, but also the variance of the random effect Ψ_{0i}, denoted as $\sigma^2_{\Psi_0}$. However, Latent GOLD uses a somewhat different (factor-analytic) parameterization of the same random-intercept model; that is,

$$\eta_{\mathbf{z}_{it},F_{1i}} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z^{pred}_{itq} + \lambda_{01} \cdot F_{1i}. \tag{35}$$

[50] Random-effects models are also referred to as multilevel, hierarchical, mixed-effects, mixed, and random-coefficients models.

On the one hand, the variance of latent variable F_{1i} is fixed to 1, but on the other hand it has an effect parameter λ_{01} associated with it. The connection between the two parameterizations is rather simple: $\sigma^2_{\Psi_0} = (\lambda_{01})^2$, which means that λ_{01} can be interpreted as the standard deviation of the random intercept.

For ordinal and nominal responses, the random intercept model has the following form:

$$\begin{aligned}
\eta_{m,\mathbf{z}_{it},F_{1i}} &= \beta_{m0} + \sum_{q=1}^{Q} \beta_{\cdot q} \cdot y^*_m \cdot z^{pred}_{itq} + \lambda_{01} \cdot y^*_m \cdot F_{1i}, \\
\eta_{m,\mathbf{z}_{it},F_{1i}} &= \beta_{m0} + \sum_{q=1}^{Q} \beta_{mq} \cdot z^{pred}_{itq} + \lambda_{m01} \cdot F_{1i}.
\end{aligned}$$

Note that for nominal variables, a single CFactor is used to capture the variation in each of the $M-1$ intercept terms, yielding the random-effects multinomial logistic regression model proposed by Hedeker (2003).

The models described above can be expanded to include *random slopes* in addition to a random intercept. However, a slight complication is that one has to decide whether the multiple random effects should be uncorrelated or not. For uncorrelated random effects, expanding the model of equation (35) with a random slope for the first predictor yields

$$\eta_{\mathbf{z}_{it},\mathbf{F}_i} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z^{pred}_{itq} + \lambda_{01} \cdot F_{1i} + \lambda_{12} \cdot F_{2i} \cdot z^{pred}_{it1}.$$

The variances of the random intercept and the random slope for z^{pred}_{it1} equal $\sigma^2_{\Psi_0} = (\lambda_{01})^2$ and $\sigma^2_{\Psi_1} = (\lambda_{12})^2$, respectively.

The same model, but now with correlated random effects can be defined as follows:

$$\eta_{\mathbf{z}_{it},\mathbf{F}_i} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z^{pred}_{itq} + \lambda_{01} \cdot F_{1i} + \lambda_{02} \cdot F_{2i} + \lambda_{12} \cdot F_{2i} \cdot z^{pred}_{it1}.$$

As can be seen, here F_{2i} does not only affect the effect of z^{pred}_{it1}, but also the intercept. Now, $\mathbf{\Sigma_\Psi}$, the variance-covariance matrix of the random effects, has to be obtained by $\mathbf{\Sigma_\Psi} = \mathbf{\Lambda}\,\mathbf{\Lambda}'$, where $\mathbf{\Lambda}$ is a matrix collecting the λ_{qd} parameters. More specifically, in our example, $\sigma^2_{\Psi_0} = (\lambda_{01})^2 + (\lambda_{02})^2$, $\sigma^2_{\Psi_1} = (\lambda_{12})^2$, and $\sigma_{\Psi_0,\Psi_1} = \lambda_{02} \cdot \lambda_{12}$.

An alternative formulation is

$$\eta_{\mathbf{z}_{it},\mathbf{F}_i} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z_{itq}^{pred} + \lambda_{01} \cdot F_{1i} + \lambda_{11} \cdot F_{1i} \cdot z_{it1}^{pred} + \lambda_{12} \cdot F_{2i} \cdot z_{it1}^{pred},$$

in which F_{1i} also affects the predictor effect. Again $\mathbf{\Sigma_\Psi} = \mathbf{\Lambda}\,\mathbf{\Lambda}'$. It should be noted that $\mathbf{\Lambda}$ is in fact the Cholesky decomposition of $\mathbf{\Sigma_\Psi}$, which is the matrix transforming $N(\mathbf{0}, \mathbf{I})$ random variables into $N(\mathbf{0}, \mathbf{\Sigma_\Psi})$ random variables; that is, $\mathbf{\Psi}_i = \mathbf{\Lambda}\,\mathbf{F}_i$.

Whereas the random-effects models presented so far contained as many CFactors as random effects, this does not mean that this is necessarily the case. In fact, with three CFactors – the Latent GOLD maximum – one can define models with any number of random effects. This is accomplished with the following "factor-analytic" specification:

$$\eta_{\mathbf{z}_{it},\mathbf{F}_i} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z_{itq}^{pred} + \sum_{d=1}^{3} \lambda_{0d} \cdot F_{di} + \sum_{d=1}^{3} \sum_{q=1}^{Q} \lambda_{qd} \cdot F_{di} \cdot z_{itq}^{pred}, \qquad (36)$$

where again $\mathbf{\Sigma_\Psi} = \mathbf{\Lambda}\,\mathbf{\Lambda}'$. This "factor-analytic" specification in which each CFactor may be associated with multiple random effects is equivalent to the generalized random coefficient (GRC) formulation proposed by Skrondal and Rabe-Hesketh (2004, p. 101). In fact, it is assumed that the unobserved heterogeneity in the regression coefficients can be summarized by three underlying CFactors.

11.2.5 Random-intercept model with covariates

Suppose you wish to regress a random intercept term on a set of covariates. This would imply that $F_{1i} = \alpha_0 + \sum_{r=1}^{R} \alpha_r \cdot z_{ir}^{cov} + \varepsilon_{1i}$. Although there is no special option in Latent GOLD for defining regression models for CFactors, substitution into the model described in equation (35) yields

$$\begin{aligned}
\eta_{\mathbf{z}_{it},F_{1i}} &= \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z_{itq}^{pred} + \lambda_{01} \cdot (\alpha_0 + \sum_{r=1}^{R} \alpha_r \cdot z_{ir}^{cov} + \varepsilon_{1i}) \\
&= \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z_{itq}^{pred} + \lambda_{01} \cdot \alpha_0 + \sum_{r=1}^{R} \lambda_{01} \cdot \alpha_r \cdot z_{ir}^{cov} + \lambda_{01} \cdot \varepsilon_{1i} \\
&= \beta_0^* + \sum_{q=1}^{Q} \beta_q \cdot z_{itq}^{pred} + \sum_{r=1}^{R} \alpha_r^* \cdot z_{ir}^{cov} + \lambda_{01} \cdot \varepsilon_{1i}.
\end{aligned}$$

This shows that regressing F_{1i} on covariates is equivalent to using the z_{ir}^{cov} as additional predictors in the regression model for the dependent variable. The original covariate effects α_r equal α_r^*/λ_{01}. Moreover, $(\lambda_{01})^2$ is no longer the total variance of F_{1i}, but the variance of the error in the regression of F_{1i} on z_{ir}^{cov}. The total variance of F_{1i} can, of course, be determined in a model without covariates.

The above specification can also be used to specify the Rasch model with covariates proposed by Zwinderman (1991). As already explained above, a Rasch model for T items can be specified as a LC Regression model in which $T-1$ item dummies are used as predictors.

11.2.6 LC (FM) regression models with random effects

The unique feature of Latent GOLD is that it allows you to combine random effects with latent classes. More specifically, it is possible to specify LC Regression models in which the intercept and/or some of the regression coefficients vary within latent classes. Lenk and DeSarbo (2000) proposed using random effects in FMs of generalized linear models, Böckenholt (2001) proposed using random effects in LC Regression models for ranking data, and Muthén (2004) and Vermunt (2007) proposed including random effects in LC growth models.

It has been observed that the solution of a LC Regression analysis may be strongly affected by heterogeneity in the intercept. In *rating-based conjoint studies*, for example, it is almost always the case that respondents differ with respect to the way they use the, say, 7-point rating scale: some respondents tend to give higher ratings than others, irrespective of the characteristics of the rated products. A LC Regression model captures this response heterogeneity phenomenon via Classes with different intercepts. However, most likely, the analyst is looking for a relatively small number of latent classes that differ in more meaningful ways with respect to predictor effects on the ratings. By including a random intercept in the LC Regression model, for example,

$$\eta_{m,x,\mathbf{z}_{it},F_{1i}} = \beta_{xm0} + \sum_{q=1}^{Q} \beta_{x.q} \cdot y_m^* \cdot z_{itq}^{pred} + \lambda_{x01} \cdot y_m^* \cdot F_{1i},$$

it is much more likely that one will succeed in finding such meaningful Classes (segments). The random intercept, which may have a different effect in each latent class, will filter out (most of) the "artificial" variation in the intercept.

Another interesting application of random effects within latent classes occurs in the context of LC *growth modeling* (Muthén, 2004; Vermunt, 2007). Suppose we have a model for a binary response variable measured at T occasions in which z_{it1}^{pred} equals time and z_{it2}^{pred} time squared. A LC growth model with a random intercept and a random slope for the linear time effect would be of the form:

$$\eta_{x,\mathbf{z}_{it},\mathbf{F}_i} = \beta_{x0} + \beta_{x1} \cdot z_{it1}^{pred} + \beta_{x2} \cdot z_{it2}^{pred} + \lambda_{01} \cdot F_{1i} + \lambda_{02} \cdot F_{2i} + \lambda_{12} \cdot F_{2i} \cdot z_{it1}^{pred},$$

where we assume that the β parameters are Class dependent and λ parameters Class independent. Similarly, LC growth models can be formulated for dependent variables of other scale types.

12 Multilevel LC Model

12.1 Model Components and Estimation Issues

To be able to explain the multilevel LC model implemented in Latent GOLD, we have to introduce and clarify some terminology. Higher-level observations will be referred to as groups and lower-level observations as cases. The records of cases belonging to the same group are connected by the *Group ID* variable. It should, however, be noted that higher-level observations can also be individuals, for example, in longitudinal applications. "Cases" would then be the multiple time points within individuals and replications (or indicators) the multiple responses of an individual at the various time points.

The index j is used to refer to a particular group and I_j to denote the number of cases in group j. With y_{jit} we denote the response on indicator (at replication) t of case i belonging to group j, with $\mathbf{y}_{ji}$ the full vector of responses of case i in group j, and with $\mathbf{y}_j$ the responses of all cases in group j. Rather than expanding the notation with new symbols, group-level quantities will be referred to using a superscript g: Group-level classes (GClasses), group-level continuous factors (GCFactors), and group-level covariates (GCovariates) are denoted by x^g, $\mathbf{F}_j^g$, and $\mathbf{z}_j^g$, and group-level parameters by γ^g, β^g, and λ^g.

The most general Latent GOLD probability structure for a multilevel LC model equals

$$f(\mathbf{y}_j|\mathbf{z}_j, \mathbf{z}_j^g) = \sum_{x^g=1}^{K^g} \int_{\mathbf{F}_j^g} f(\mathbf{F}_j^g)\, P(x^g|\mathbf{z}_j^g)\, f(\mathbf{y}_j|\mathbf{z}_j, x^g, \mathbf{F}_j^g)\, \mathrm{d}\mathbf{F}_j^g, \qquad (37)$$

where

$$f(\mathbf{y}_j|\mathbf{z}_j, x^g, \mathbf{F}_j^g) = \prod_{i=1}^{I_j} f(\mathbf{y}_{ji}|\mathbf{z}_{ji}, x^g, \mathbf{F}_j^g).$$

Assuming that the model of interest may also contain CFactors, for each case i, $f(\mathbf{y}_{ji}|\mathbf{z}_{ji}, x^g, \mathbf{F}_j^g)$ has a structure similar to the one described in equation (33); that is,

$$f(\mathbf{y}_{ji}|\mathbf{z}_{ji}, x^g, \mathbf{F}_j^g) = \sum_{x=1}^{K} \int_{\mathbf{F}_{ji}} f(\mathbf{F}_{ji})\, P(x|\mathbf{z}_{ji}, x^g, \mathbf{F}_j^g)\, f(\mathbf{y}_{ji}|x, \mathbf{z}_{ji}, \mathbf{F}_{ji}, x^g, \mathbf{F}_j^g)\, \mathrm{d}\mathbf{F}_{ji}.$$

where

$$f(\mathbf{y}_{ji}|x, \mathbf{z}_{ji}, \mathbf{F}_{ji}, x^g, \mathbf{F}_j^g) = \prod_{h=1}^{H} f(\mathbf{y}_{jih}|x, \mathbf{z}_{ji}, \mathbf{F}_{ji}, x^g, \mathbf{F}_j^g)$$

These four equations show that a multilevel LC model is a model

- for $f(\mathbf{y}_j|\mathbf{z}_j, \mathbf{z}_j^g)$, which is the marginal density of all responses in group j given all exogenous variable information in group j,
- containing GClasses (x^g) and/or (at most three mutually independent) GCFactors ($\mathbf{F}_j^g$),
- containing GCovariates $\mathbf{z}_j^g$ affecting the group classes x^g,
- assuming that the I_j observations for the cases belonging to group j are independent of one another given the GClasses and GCFactors,
- allowing the GClasses and GCFactors to affect the case-level latent classes x and/or the responses $\mathbf{y}_{ji}$.

GCFactors enter in exactly the same manner in the linear predictors for the various types of response variables as case-level CFactors. We will refer to their coefficients as $\lambda_d^{t,g}$ (Cluster and DFactor) and λ_{xqd}^g (Regression), where we add a subscript m when needed. GCFactors can also be used in the model for the latent classes. These terms are similar to those for nominal (Cluster and Regression) or ordinal (DFactor) dependent variables. We will denote a GCFactor effect on the latent classes as $\lambda_{xrd}^{0,g}$, $0 \leq 1 \leq R$, where the superscript 0 refers to the model for the latent classes.

GClasses enter in the linear predictors of the models for the indicators as $\beta_{x^g}^{t,g}$ and in the one of the model for the dependent variable as

$\beta^g_{x0,x^g} + \sum_{q=1}^{Q} \beta^g_{xq,x^g} \cdot z^{pred}_{jitq}$. Inclusion of GClasses in the model for the Clusters, DFactors, or Classes implies that the γ parameters become GClass dependent; that is $\eta_{x|\mathbf{z}_{ji},x^g} = \gamma_{x^g,x0} + \sum_{r=1}^{R} \gamma_{x^g,xr} \cdot z^{cov}_{jir}$. Note that this is similar to a LC Regression analysis, where x^g now plays the role of x, and x the role of a nominal or ordinal y variable.

The remaining linear predictor is the one appearing in the multinomial logistic regression model for the GClasses. It has the form $\eta_{x^g|\mathbf{z}^g_i} = \gamma^g_{x^g,0} + \sum_{r=1}^{R^g} \gamma^g_{x^g,r} \cdot z^{g,cov}_{jr}$. This linear predictor is similar to the one for the Clusters or Classes (in a standard LC model), showing that GCovariates may be allowed to affect GClasses in the same way that covariates may affect Classes.

Below we will describe the most relevant special cases of this very general latent variable model,[51] most of which were described in Vermunt (2002b, 2003, 2004, and 2005) and Vermunt and Magidson (2005b). We then devote more attention to the expressions for the exact forms of the various linear predictors in models with GClasses, GCFactors, and GCovariates.

Model restrictions Both in Cluster and in DFactor models, one can equate the λ's are across indicators of the same type for selected CFactors ("Equal Effects") and/or fix some of the λ's to zero. The same applies to the β's corresponding to the GClasses.

In Regression, one can use the parameter constraints "Class Independent", "No Effect", and "Merge Effects", implying equal λ's (β's) among all Classes, zero λ's (β's) in selected Classes, and equal λ's (β's) in selected Classes.

ML (PM) estimation and technical settings Similar to what was discussed in the context of CFactors, with GCFactors, the marginal density $f(\mathbf{y}_j|\mathbf{z}_j)$ described in equation (37) is approximated using Gauss-Hermite quadrature. With three GCFactors and B quadrature nodes per dimension, the approximate density equals

$$f(\mathbf{y}_j|\mathbf{z}_j, \mathbf{z}^g_j) \approx \sum_{x^g=1}^{K^g} \sum_{b_1=1}^{B} \sum_{b_2=1}^{B} \sum_{b_3=1}^{B} P(x^g|\mathbf{z}^g_j)\, f(\mathbf{y}_j|\mathbf{z}_j, x^g, F^g_{b_1}, F^g_{b_2}, F^g_{b_3})\, P^g_{b_1}\, P^g_{b_2}\, P^g_{b_3}.$$

[51] In fact, the multilevel LC model implemented in Latent GOLD is so general that many possibilities remain unexplored as of this date. It is up to Latent GOLD Advanced users to further explore its possibilities.

ML (PM) estimates are found by a combination of the upward-downward variant of the EM algorithm developed by Vermunt (2003, 2004, 2008a) and Newton-Raphson. The first- and second-order derivatives for the Newton-Raphson iterations are obtained analytically using a recursion scheme similar to the one described by Lystig and Hughes (2002) for latent Markov models.

The only additional technical setting in multilevel LC models is the same one as in models with CFactors; that is, the setting related to the number of quadrature nodes to be used in the numerical integration concerning the GCFactors. As already explained in the context of models with CFactors, the default value is 10, the minimum 2, and the maximum 50.

12.2 Application Types

12.2.1 Two-level LC or FM model

The original multilevel LC model described by Vermunt (2003) and Vermunt and Magidson (2005b) was meant as a tool for multiple-group LC analysis in situations in which the number of groups is large. The basic idea was to formulate a model in which the latent class distribution (class sizes) is allowed to differ between groups by using a random-effects approach rather than by estimating a separate set of class sizes for each group – as is done in a traditional multiple-group analysis.

When adopting a nonparametric random-effects approach (using GClasses), one obtains the following multilevel LC model:

$$f(\mathbf{y}_j) = \sum_{x^g=1}^{K^g} P(x^g) \left[\prod_{i=1}^{I_j} \sum_{x=1}^{K} P(x|x^g) \prod_{t=1}^{T} f(y_{jit}|x) \right],$$

in which the linear predictor in the logistic model for $P(x|x^g)$ equals $\eta_{x|x^g} = \gamma_{x^g,x0}$. Here, we are in fact assuming that the intercept of the model for the latent classes differs across GClasses.

When adopting a parametric random-effects approach (GCFactors), one obtains

$$f(\mathbf{y}_j) = \int_{-\infty}^{\infty} f(F_{1j}^g) \left[\prod_{i=1}^{I_j} \sum_{x=1}^{K} P(x|F_{1j}^g) \prod_{t=1}^{T} f(y_{jit}|x) \right] \mathrm{d}F_{1j}^g,$$

where the linear term in the model for $P(x|F_{1j}^g)$ equals $\eta_{x|F_{1j}^g} = \gamma_{x0} + \lambda_{x01}^{0,g} \cdot F_{1j}^g$. Note that this specification is the same as in a random-intercept model for a nominal dependent variable.

Vermunt (2005) expanded the above parametric approach with covariates and random slopes, yielding a standard random-effects multinomial logistic regression model, but now for a *latent* categorical outcome variable. With covariates and multiple random effects, we obtain

$$f(\mathbf{y}_j|\mathbf{z}_j) = \int_{\mathbf{F}_j^g} f(\mathbf{F}_j^g) \left[\prod_{i=1}^{I_j} \sum_{x=1}^{K} P(x|\mathbf{z}_{ji}, \mathbf{F}_j^g) \prod_{t=1}^{T} f(y_{jit}|x) \right] \mathrm{d}\mathbf{F}_j^g,$$

where the linear predictor for x equals

$$\eta_{x|\mathbf{z}_{ji},\mathbf{F}_j^g} = \gamma_{x0} + \sum_{r=1}^{R} \gamma_{xr} \cdot z_{jir}^{cov} + \sum_{d=1}^{D^g} \lambda_{x0d}^{0,g} \cdot F_{dj}^g + \sum_{d=1}^{D^g} \sum_{r=1}^{R} \lambda_{xrd}^{0,g} \cdot F_{dj}^g \cdot z_{jir}^{cov},$$

Whereas in the Cluster and Regression submodules, this is a random-effects multinomial logistic regression model, in the DFactor submodule, we use a random-effects ordinal logistic regression model for each of the discrete ordinal factors.

Also when adopting a nonparametric random-effects approach, one may include covariates in the multilevel LC model; that is,

$$\eta_{x|\mathbf{z}_{ji},x^g} = \gamma_{x^g,x0} + \sum_{r=1}^{R} \gamma_{x^g,xr} \cdot z_{jir}^{cov}.$$

This yields a model for the latent classes in which the intercept and the covariate effects may differ across GClasses. In fact, we have a kind of LC Regression structure in which the latent classes serve as a nominal dependent variable and the GClasses as latent classes.

An important extension of the above nonparametric multilevel LC models is the possibility to regress the GClasses on group-level covariates. This part of the model has the same form as the multinomial logistic regression model for the Clusters or Classes in a standard LC or FM model.

In the Cluster and DFactor submodules, it is possible to allow GCFactors and/or GClasses to have direct effects on the indicators. As suggested by Vermunt (2003), this is a way to deal with item bias. Below, we will discuss various other applications of this option.

12.2.2 LC (FM) regression models for three-level data

Another application type of the Latent GOLD multilevel LC option is three-level regression modeling (Vermunt, 2004). This application type concerns the Regression submodule.

A three-level LC (FM) Regression model would be of the form

$$f(\mathbf{y}_j|\mathbf{z}_j) = \sum_{x^g=1}^{K^g} P(x^g) \left[\prod_{i=1}^{I_j} \sum_{x=1}^{K} P(x) \prod_{t=1}^{T_i} f(y_{jit}|x, \mathbf{z}_{jit}^{pred}, x^g) \right].$$

Suppose we have a LC Regression model for a binary outcome variable. The simplest linear predictor in a model that includes GClasses would then be

$$\eta_{\mathbf{z}_{jit},x,x^g} = \beta_{x0} + \sum_{q=1}^{Q} \beta_{xq} \cdot z_{jitq}^{pred} + \beta_{0,x^g}^{g},$$

which is a model in which (only) the intercept is affected by the GClasses. A more extended model is obtained by assuming that also the predictor effects vary across GClasses; that is,

$$\eta_{\mathbf{z}_{jit},x,x^g} = \beta_{x0} + \sum_{q=1}^{Q} \beta_{xq} \cdot z_{jitq}^{pred} + \beta_{0,x^g}^{g} + \sum_{q=1}^{Q} \beta_{q,x^g}^{g} \cdot z_{jitq}^{pred}.$$

In practice, it seems to be most natural to allow effects of predictors that change values across replications to be Class dependent and effects of predictors that change values across cases to depend on the GClasses.

The most extended specification is obtained if all the effects are assumed to be Class dependent, which implies including Classes-GClasses (x-x^g) interactions. Such a model is defined as

$$\eta_{\mathbf{z}_{jit},x,x^g} = \beta_{x0} + \sum_{q=1}^{Q} \beta_{xq} \cdot z_{jitq}^{pred} + \beta_{x0,x^g}^{g} + \sum_{q=1}^{Q} \beta_{xq,x^g}^{g} \cdot z_{jitq}^{pred}.$$

It should be noted that in each of the above three models, identifying constraints have to be imposed on the parameters involving the GClasses. In the most general model, this is $\sum_{x^g=1}^{K^g} \beta_{xq,x^g}^{g} = 0$, $\beta_{xq1}^{g} = 0$, or $\beta_{xqK^g}^{g} = 0$, for $0 \leq q \leq Q$ and $1 \leq x \leq K$. In other words, the parameters in the model for the dependent variable either sum to zero across GClasses, are equal to zero for the first GClass, or are equal to zero for the last GClass.

12.2.3 Three-level random-effects GLMs

Combining the GCFactors from the multilevel model with the CFactors option makes it possible to specify "standard" three-level GLM regression models with parametric random effects (Im and Gionala, 1988; Skrondal and

Rabe-Hesketh, 2004; Rodriguez and Goldman, 2001; Vermunt, 2002c, 2004). In terms of probability structure, this yields:

$$f(\mathbf{y}_j|\mathbf{z}_j) = \int_{\mathbf{F}_j^g} f(\mathbf{F}_j^g) \left[\prod_{i=1}^{I_j} \int_{\mathbf{F}_{ji}} f(\mathbf{F}_{ji}) \prod_{t=1}^{T_i} f(y_{jit}|\mathbf{z}_{jit}^{pred}, \mathbf{F}_{ji}, \mathbf{F}_j^g) \mathrm{d}\mathbf{F}_{ji} \right] \mathrm{d}\mathbf{F}_j^g.$$

The simplest special case is obtained by assuming that the regression model contains random intercepts at both the case and the group level. Taking again a binary logistic regression model as an example, the corresponding linear predictor equals

$$\eta_{\mathbf{z}_{jit}, F_{1ji}, F_{1j}^g} = \beta_0 + \sum_{q=1}^{Q} \beta_q \cdot z_{jitq}^{pred} + \lambda_{01} \cdot F_{1ji} + \lambda_{01}^g \cdot F_{1j}^g.$$

Such a model containing a single CFactor and a single GCFactor will suffice in most three-level random-effects applications. However, similarly to the random effects models discussed in the context of the CFactors option, this model can be expanded with random slopes at both levels using the factor-analytic random-effects specification illustrated in equation (36).

12.2.4 LC growth models for multiple indicators or nested data

Suppose one has a longitudinal data set containing multiple indicators (response variables) for each time point. The multiple responses could be used to build a time-specific latent classification, while the pattern of (latent) change over time could be described using a (LC) growth model. Specification of such a model would involve using the index i for the time points and the index j for the cases (time points are nested within cases). The LC model for the time points may have the form of a LC Cluster model, but also of one of the other types of Latent GOLD models. The multinomial logistic regression model for the (time-specific) latent classes will have the form of a LC growth model: class membership depends on time, where the intercept and possibly also the time slope is allowed to vary across individuals. This variation can be modelled using continuous random effects (GCFactors) and/or discrete random effects (GClasses).

Palardy and Vermunt (2009) describe a latent growth model for multilevel data sets. Higher-level units (schools) are assumed to belong to group-level latent classes which differ in the growth trajectories of their pupils. The growth model for the pupils contains a random intercept and a random slope.

12.2.5 Various IRT applications

Several authors have proposed multilevel extensions of IRT models (Fox and Glas, 2001; Skrondal and Rabe-Hesketh, 2004). A possible specification of a *multilevel IRT model* is

$$\eta^t_{F_{ji},F^g_j} = \beta^t_0 + \lambda^t \cdot F_{ji} + \lambda^{t,g} \cdot F^g_j,$$

which besides the lower-level trait F_{ji} contains a higher-level (group-level) trait or random effect F^g_j. This is an example of a two-level two-parameter logistic model. A two-level Rasch model would be of the form

$$\eta^t_{F_{ji},F^g_j} = \beta^t_0 + \lambda \cdot F_{ji} + \lambda^g \cdot F^g_j.$$

Tay, Diener, Drasgow, and Vermunt (2011), Varriale and Vermunt (2012), and Vermunt (2008b) describe multilevel IRT models (and factor models for categorical responses) with group-level classes.

A strongly related type of IRT model that can be defined with the Latent GOLD multilevel option is the *state-trait model* for ordinal variables described by Steyer and Partchev (2001). This is, in fact, a two-level IRT model for repeated measures data (i indexes occasions and j cases). Assuming that there are two underlying traits, their state-trait model for ordinal responses has the following form:

$$\eta^t_{m,F_{ji},F^g_j} = \beta^t_{m0} + \lambda^t_{\cdot} \cdot y^{*t}_m \cdot F_{ji} + \lambda^{t,g}_{\cdot 1} \cdot y^{*t}_m \cdot F^g_{1j} + \lambda^{t,g}_{\cdot 2} \cdot y^{*t}_m \cdot F^g_{2j},$$

where F_{ji} is the time-specific latent state variable and F^g_{1j} and F^g_{2j} are two time-constant latent trait variables. This model is, in fact, a two-level variant of the generalized partial-credit model. More restricted models can be defined by assuming, as in a partial-credit model, that the λ parameters are equal across items and/or, as in a rating-scale model, that the intercept has form $\beta^t_{m0} = \beta^{\cdot}_{m0} + \beta^t_{\cdot 0} \cdot y^{*t}_m$.

In the state-trait model, the latent state variable captures local dependencies between the responses at the same occasion. The same type of structure can be used to deal with other types of situations in which item belonging to the same set cannot be assumed to be locally independent, such as items belonging to the same booklet of a test.

12.2.6 Non multilevel models

The last possible use of the multilevel option we would like to mention does not yield a multilevel model, but is a trick for estimating models that cannot be estimated in another way. The trick consists of using a Group ID variable that is identical to the Case ID or, equivalently, to have groups consisting of no more than one case. GCFactors can then be used as CFactors. This makes it possible to define models in which CFactors – for example, continuous traits or factors measured by multiple indicators – affect the latent classes. Another possibility is to use the GClasses as an additional case-level nominal latent variable, yielding a model in which one nominal latent variable may affect another nominal latent variable.

13 Complex Survey Sampling

The Survey option makes it possible to obtain consistent parameter estimates and correct standard errors with complex sampling designs. The option can be used in combination with any model that can be estimated with Latent GOLD. Parameter estimation is based on the so-called pseudo-ML estimator that uses the sampling weights as if it were case weights. Correct statistical tests with stratified and clustered samples, as well as with sampling weights and samples from finite populations are obtained using the linearization variance estimator.

Latent GOLD also implement an alternative method to deal with sampling weights. This is a two-step procedure in which the model is first estimated without making use of the sampling weights, and in which subsequently the latent class sizes and covariate effects are corrected using the sampling weights.

13.1 Pseudo-ML Estimation and Linearization Estimator

The survey option can be used to take into account the fact that cases may

1. belong to the same stratum,
2. belong to the same primary sampling unit (PSU), often referred to as a sampling cluster,

3. contain a sampling weight,

4. be sampled from a finite population.

Let o denote a particular stratum, c a particular PSU in stratum o, and i a particular case in PSU c of stratum o. Moreover, let O be the number of strata, C_o the number of PSUs in stratum o, and I_{oc} the number of cases in PSU c of stratum o. The sampling weight corresponding to case i belonging to PSU c of stratum o is denoted by sw_{oci}, and the population size (total number of PSUs) of stratum o by N_o.[52]

From this notation, it can be seen that PSUs are nested within strata, and that cases are nested within PSUs. In other words, records with the same Case ID should belong to the same PSU, and all records with the same PSU identifier should belong to the same stratum. The population size N_o indicates the population number of PSUs in stratum o, and should thus have the same value across records belonging to the same stratum. Another thing that should be noted is that in *multilevel models*, the strata, PSUs, and sampling weights concern groups rather than cases; that is, one has strata and PSUs formed by groups and sampling weights for groups.

For parameter estimation, only the sampling weights need to be taken into account. When sampling weights are specified, Latent GOLD will estimate the model parameters by means of pseudo-ML (PM) estimation (Skinner, Holt, and Smith, 1989). Recall that ML estimation involves maximizing

$$\log \mathcal{L} = \sum_{i=1}^{I} w_i \log f(\mathbf{y}_i|\mathbf{z}_i, \boldsymbol{\vartheta}),$$

where w_i is a case weight. In pseudo-ML estimation, one maximizes

$$\log \mathcal{L}_{pseudo} = \sum_{o=1}^{O} \sum_{c=1}^{C_c} \sum_{i=1}^{I_{oc}} sw_{oci} \log f(\mathbf{y}_{oic}|\mathbf{z}_{oic}, \boldsymbol{\vartheta}),$$

which is equivalent to maximizing $\log \mathcal{L}$ using the sampling weights as if it were case weights. In Latent GOLD one may also have both case and sampling weights, in which case we get

$$\log \mathcal{L}_{pseudo} = \sum_{o=1}^{O} \sum_{c=1}^{C_c} \sum_{i=1}^{I_{oc}} w_{oci} \cdot \left[sw_{oci} \log f(\mathbf{y}_{oic}|\mathbf{z}_{oic}, \boldsymbol{\vartheta}) \right],$$

[52] In Latent GOLD, one can either specify the fraction $\frac{C_o}{N_o}$ or the population size N_o. If the specified number in "Population Size" is smaller than 1 it is interpreted as a fraction, otherwise as a population size.

which is equivalent to performing ML estimation using the $sw_{oci} \cdot w_{oci}$ as "case" weights.

Each of the four complex sampling characteristics is taken into account by the so-called linearization estimator of variance-covariance matrix of the parameter estimates (Skinner, Holt, and Smith, 1989). Application of this method in the context of FM and LC models was proposed by Vermunt (2002b) and Wedel, Ter Hofstede, and Steenkamp (1998). The overall structure of $\widehat{\Sigma}_{survey}(\boldsymbol{\vartheta})$ is similar to the robust or sandwich estimator $\widehat{\Sigma}_{robust}(\boldsymbol{\vartheta})$ discussed earlier; that is,

$$\widehat{\Sigma}_{survey}(\boldsymbol{\vartheta}) = \widehat{\mathbf{H}}^{-1}\,\widehat{\mathbf{B}}\;\widehat{\mathbf{H}}^{-1}.$$

As can be seen, a matrix $\mathbf{B}$ is "sandwiched" between the inverse of the Hessian matrix. For the computation of $\mathbf{B}$, one needs two components: the contribution of PSU c in stratum o to the gradient of parameter k, denoted by g_{ock}, and its sample mean in stratum o, denoted by $\overline{g}_{ok}$. These are obtained as follows:

$$g_{ock} = \sum_{i=1}^{I_{oc}} sw_{oci} \frac{\partial \log f(\mathbf{y}_{oci}|\mathbf{z}_{oci}, \boldsymbol{\vartheta})}{\partial \vartheta_k}$$

and

$$\overline{g}_{ok} = \frac{\sum_{c=1}^{C_o} g_{ock}}{C_o}$$

Using these two components, element $B_{kk'}$ of $\mathbf{B}$ can be defined as

$$B_{kk'} = \sum_{o=1}^{O} \frac{C_o}{C_o - 1}(1 - \frac{C_o}{N_o}) \sum_{c=1}^{C_o}(g_{ock} - \overline{g}_{ok})(g_{ock'} - \overline{g}_{ok'}).$$

Note that if we neglect the finite population correction factor $(1-\frac{C_o}{N_o})$, $\mathbf{B}$ is the sample covariance matrix of the PSU-specific contributions to the gradient vector.

Various observations can be made from the formula for $B_{kk'}$. The first is that without complex sampling features (one stratum, single case per PSU, no sampling weights, and $\frac{C_o}{N_o} \approx 0$), the above procedure yields $\widehat{\boldsymbol{\Sigma}}_{robust}(\boldsymbol{\vartheta})$, which shows that $\widehat{\boldsymbol{\Sigma}}_{survey}(\boldsymbol{\vartheta})$ does not only take into account the sampling design, but is also a robust estimator of $\boldsymbol{\Sigma}(\boldsymbol{\vartheta})$. Second, the fact that gradient contributions are aggregated for cases belonging to the same PSU shows that the PSUs are treated as the independent observational units, which is exactly what we want. Third, the term $\frac{C_o}{C_o-1}$ is only defined if each stratum contains

at least two PSUs: Latent GOLD "solves" this problem by skipping strata for which $C_o = 1$, which implies the single PSU has a contribution of 0 to the covariance of the gradients. A common solution to this problem is to merge strata.[53]

The design effect corresponding to a single parameter equals the ratio of its design corrected variance and its variance assuming simple random sampling. A multivariate generalization is obtained as follows (Skinner, Holt, and Smith, 1989):

$$\begin{aligned} deff &= \text{tr}\left[\widehat{\boldsymbol{\Sigma}}_{standard}(\boldsymbol{\vartheta})^{-1}\widehat{\boldsymbol{\Sigma}}_{survey}(\boldsymbol{\vartheta})\right] / npar \\ &= \text{tr}\left[(-\widehat{\mathbf{H}})\,\widehat{\mathbf{H}}^{-1}\,\widehat{\mathbf{B}}\;\widehat{\mathbf{H}}^{-1}\right] / npar = \text{tr}\left[-\widehat{\mathbf{B}}\;\widehat{\mathbf{H}}^{-1}\right] / npar, \end{aligned}$$

where "tr" is the trace operator. The generalized design effect is thus the average of the diagonal elements of $-\widehat{\mathbf{B}}\,\widehat{\mathbf{H}}^{-1}$. Note that this number equals the average of the eigenvalues of this matrix.

13.2 A Two-step Method

Latent GOLD also implements an alternative two-step method for dealing with sampling weights in LC analysis, which was described in Vermunt and Magidson (2001) and Vermunt (2002b). The procedure involves performing an unweighted analysis followed by a weighted analysis in which the parameters in the model part for the response variables are fixed to their unweighted ML (PM) estimates. More specifically, in step two, the class sizes and the covariates effects are adjusted for the sampling weights. The adjusted log-likelihood function that is maximized equals

$$\log \mathcal{L}_{adjusted} = \sum_{i=1}^{I} sw_i \log \sum_{x=1}^{K} P(x|\mathbf{z}_i, \boldsymbol{\vartheta}_{adjusted})\, f(\mathbf{y}_i|x, \mathbf{z}_i, \widehat{\boldsymbol{\vartheta}}_{ML}),$$

where $\boldsymbol{\vartheta}_{adjusted}$ are the unknown parameters to be estimated.

The rationale of this procedure is that an unweighted analysis may yield more stable (more efficient) estimates for the parameters defining the latent classes, but yields biased class sizes and covariate effects. The latter are corrected in the second step of the procedure.

[53] LG-Syntax implements two other ways of dealing with this problem.

14 Latent GOLD's Advanced Output

This section describes the changes and additional items in the Latent GOLD output sections when the Advanced options are used.

14.1 Model Summary

For Markov models, the first part of the Model Summary output reports the total number time points (aggregated over all cases; $\sum_i^I w_i T_i$) in addition to the number of cases. For multilevel models, it also reports the number of groups (J). When the Survey option is used, the program reports the generalized design effect ($deff$), which is an overall measure indicating how many times larger the design corrected variances are compared to the asymptotic variances.

For multilevel models, *Chi-squared Statistics* are not reported and the bootstrap L^2 and $-2LL$-difference options are not available (the latter is available in Syntax). When the Survey option is used, the bootstrap-based L^2 and LL-difference tests are corrected for the complex sampling design by multiplying the bootstrap replications' L^2 and $-2LL$-difference values by the generalized design effect $deff$. Note that the bootstrap replication samples themselves are obtained by simple random sampling.

In multilevel models, as in all other Latent GOLD models, the number of cases serves as N (sample size) in the computation of the BIC, CAIC, and SABIC values that appear in the *Log-likelihood Statistics*. However, as shown by Lukociene and Vermunt (2010), and Lukociene, Varriale, and Vermunt (2010), in multilevel LC models it is better to use the number of groups in the information criteria formulae. Therefore, we also provide versions of BIC, CAIC and SABIC based on the log-likelihood using $Ngroups$ ($=J$) as the sample size.

The *Classification Statistics* contain information on how well one can predict an individual's State membership and CFactor scores and a group's GClass membership and GCFactor scores. For States and GClasses, one obtains the same information as for the latent classes (proportion of classification errors and three R^2 measures). For CFactors and GCFactors, one obtains only the standard R^2, which can be interpreted as a reliability measure. In Markov and multilevel models with covariates, *Model Classification Statistics* will contain information for the States and GClasses. The program reports a *Classification Table* for each discrete latent variable in the

model, thus also for the States and Classes in Markov models and the group-level Classes in multilevel models. Tables for both modal and proportional classification are given.

The computation of *Entropy* is adapted for Markov and multilevel LC models. It is obtained by adding up the separate entropy values for the GClasses, the Classes, Clusters, or DFactors, and the States, which means that we ignore the dependencies between the latent variables belong to different levels. CL, CLC, AWE, and $ICL\text{-}BIC$ are computed using this definition of the total entropy.

The *Prediction Statistics* are the same as in models without CFactors, GClasses, and GCFactors. The R^2_y measures indicates how well a model predicts the dependent variable given all predictors, covariates, and latent variables.

14.2 Parameters

This section reports the parameters corresponding to CFactors, GClasses, and GCFactors. In Cluster and DFactor models, CFactors, GClasses, and GCFactors can be included in the *Models for Indicators*. In Regression models, CFactors, GClasses, and GCFactors effects may appear in the *Model for Dependent*.

In multilevel models, GClasses and GCFactors may be used in the *Model for Clusters, DFactors, or Classes*. When GClasses affect a particular term (the intercept or a covariate effect), one obtains a separate set of coefficients for each GClass. GCFactors enter as random effects in the regression model for the discrete latent variable(s).

In models with GClasses, the parameters output contains the coefficients of the multinomial logistic regression *Model for GClasses*.

The reported Class-specific $R^2_{y|x}$ measures in Regression are obtained by averaging the predicted values over the other latent variables included in the model. This is the reason that in a one-Class model, the "Class-specific" $R^2_{y|1}$ may be lower that the overall R^2_y.

When the Survey option is used, one obtains design corrected standard errors and Wald statistics.

In Cluster and DFactor models, CFactors, GClasses, and GCFactors appear in the subsection *Loadings*. For continuous indicators and continuous latent variables, these are standardized regression coefficients. In all other

cases, these are linear approximations as described in Magidson and Vermunt (2004) and Vermunt and Magidson (2005a).

In Regression models with CFactors, one obtains an output subsection called *Random Effects*. This subsection provides the CFactor effects $\mathbf{\Lambda}$ and the variance-covariance matrix of the random effects, $\mathbf{\Sigma_\Psi} = \mathbf{\Lambda}\,\mathbf{\Lambda}'$.

In Markov models, the program reports the parameters for the models for the classes (if Classes>1), the initial state, the transitions (by default using transition coding), and the indicators. The formatting of this output is somewhat different from the other GUI submodules, because internally we use the Latent GOLD Syntax system to estimate Markov models. The Parameters output formatting is therefore the same as in Syntax models.

14.3 Profile

The *Profile* output section reports exactly the same information as in non-advanced models. In models in which CFactors, GClasses, and/or GCFactors have a direct effect on one or more indicators, the *Marginal Profile* is obtained by aggregating over these latent variables (integrating over CFactors and GCFactors, and summing over GClasses). It such situations, the numbers in the *Partial Profile* are obtained at the value 0 for the CFactors and GCFactors, and – as for other nominal variables having direct effects on the indicators – at the (unweighted) average of the GClasses coefficients concerned.

For Markov models, in addition to the Class sizes and the indicator information conditional of Classes and States, Profile reports the average latent State proportions across measurement occasions (marginalized over classes and covariates), the average State probabilities conditional on Class (marginalized over covariates), and the latent transition probabilities (marginalized over classes and covariates). The Partial Profile output is not available in Markov models. Moreover, in Markov models, Profile does not contain information on Covariates.

14.4 GProfile

The first part of this output section reports the sizes of the GClasses $[P(x^g)]$ and the probability of being in a certain latent class or DFactor level for each GClass $[P(x|x^g)$ or $P(x_\ell|x^g)]$.

The second part of the GProfile section reports the GClass-specific means and probabilities for the response variables. The computation of this part

of the GProfile output is similar to the computation of the same kinds of numbers in the Profile output.

14.5 Profile-Longitudinal

For Markov models, Latent GOLD Advanced 5.1 implements a special type of Profile output and an accompanying plot called *Profile-Longitudinal*. It can be used to investigate the time trend in the indicators, per Class, per State, overall, and observed; that is, it reports the time-specific estimated and observed probabilities/means for the indicators. The computations to obtain the reported probabilities and means are the same as for (marginal) Profile, with the difference that this is done separately for each time point.

14.6 ProbMeans

In models with CFactors, the Probmeans output reports the average CFactor posterior means for categories of the indicators, the dependent variable, and covariates. In Markov models, ProbMeans reports information on both Classes and States.

14.7 Two-level and Longitudinal Bivariate Residuals

Latent GOLD 5.1 Advanced/Syntax implements two types of BVRs called two-level BVRs and longitudinal BVRs. Two-level BVRs are reported in multilevel Cluster and DFactor models, and longitudinal BVRs in Markov models.

The two types of two-level BVRs reported in multilevel Cluster and DFactor models are BVR-group and BVR-pairs. These quantify how well the multilevel model picks up between-group differences and within-group similarities in responses on the indicator concerned (Nagelkerke, Oberski, and Vermunt, in press). BVR-group is equivalent to the BVR obtained by using the groupid variable also as a nominal covariate (with its effect set equal to 0). The BVR-pairs computation for categorical indicators involves setting up the two-way cross-tabulation for the responses of pairs of observations within groups. The estimated frequencies $E(n_{m,m'})$ are obtained as follows:

$$E(n_{m,m'}) = \sum_{j=1}^{J} \sum_{i=1}^{I_j} \sum_{i'<i} w_i w_{i'} \sum_{x^g=1}^{K^g} \widehat{P}(y_{jit} = m|x^g) \widehat{P}(y_{ji't} = m'|x^g) \widehat{P}(x^g|\mathbf{z}_j, \mathbf{y}_j),$$

with

$$\widehat{P}(y_{jit} = m|x^g) = \sum_{x=1}^{K} \widehat{P}(y_{jit} = m|x, x^g)\widehat{P}(x|x^g).$$

The tables with observed and estimated frequencies are made symmetric since the order of the observations within pairs is arbitrary: the new off-diagonal entries become $(n_{m,m'} + n_{m',m})/2$ and $(E(n_{m,m'}) + E(n_{m',m}))/2$, respectively. Then, the table with estimated frequencies is adjusted using a series of IPF cycles to equate its univariate margins to those of the observed table. Because of the symmetry within pairs, in the Pearson chi-squared computation, symmetric off-diagonal cells are collapsed. BVR-pairs equals the resulting chi-squared value divided by $M \cdot (M-1)/2$ (the number of parameters of a symmetric association) and by the average group size.

The three types of longitudinal BVRs are BVR-time, BVR-lag1, and BVR-lag2. These indicate whether the estimated Markov model captures the time trend, the first-order autocorrelation, and the second-order autocorrelation, respectively, for the indicator concerned. BVR-time is equivalent to the BVR obtained by using the time variable as a nominal covariate (possibly with its effects set equal to 0). For categorical indicators, BVR-lag1 and BVR-lag2 are obtained by cross-tabulating the responses at time points $t-1$ and t (BVR-lag1) and at $t-2$ and t (BVR-lag2). For BVR-lag1,

$$E(n_{m,m'}) = \sum_{i=1}^{I} w_i \sum_{t=2}^{T_i} \sum_{x_{t-1}^d=1}^{K^d} \sum_{x_t^d=1}^{K^d} \widehat{P}(y_{it-1} = m|x_{t-1}^d)\widehat{P}(y_{it'} = m'|x^d)\widehat{P}(x_{t-1}^d, x_t^d|\mathbf{z}_j).$$

Again the margins of the estimated table are adjusted to be equal those of the observed table. BVR-lag1 equals the Pearson chi-squared value for this table divided by $(M-1)^2$ and by the average number of lag-1 responses per individual. BVR-lag2 is computed in a similar way.

As for standard BVRs, with ordinal variables, the two-level and longitudinal BVRs are computed using after adjusting the observed frequencies in the table concerned to be in agreement with an ordinal association structure. For other scale types, the two-level and longitudinal BVRs these are scaled unexplained variances (BVR-group and BVR-time) and scaled squared differences between observed and estimated correlations (BVR-pairs, BVR-lag1, BVR-lag2).

14.8 Frequencies

Frequencies are not reported in multilevel LC models.

14.9 Estimated Values

In Regression models with CFactors, GClasses, and/or GCFactors, the reported estimated values are obtained by aggregating over these latent variables (integrating over CFactors and GCFactors, and summing over GClasses). As far as these latent variables is concerned, the reported estimated values can be interpreted as model predictions.

In Cluster and DFactor models with CFactors and/or GCFactors, the reported estimated values are obtained by computing the model probabilities for CFactors and GCFactors values equal to 0 (rather than integrating over these continuous latent variables). In Markov models, Estimated Values contains the four sets of model probabilities/means appearing in equation (32).

14.10 Classification

The *Standard Classification* output provides information on the CFactor and GCFactor posterior means $\widehat{E}(F_{di}|\mathbf{z}_i, \mathbf{y}_i)$ and $\widehat{E}(F^g_{dj}|\mathbf{z}_j, \mathbf{y}_j)$, the GClass posterior probabilities $\widehat{P}(x^g|\mathbf{z}_j, \mathbf{y}_j)$ and the modal GClass for each data pattern, and the States posterior probabilities $\widehat{P}(x^d|\mathbf{z}_j, \mathbf{y}_j)$ and the modal State for each measurement occasion.

The posterior means for the continuous latent variables are obtained using Gauss-Hermite quadrature; for example,

$$\begin{aligned}\widehat{E}(F_{di}|\mathbf{z}_i, \mathbf{y}_i) &= \frac{\int_{-\infty}^{\infty} F_{di}\, f(F_{di})\, f(\mathbf{y}_i|\mathbf{z}_i, F_{di})\, \mathrm{d}\, F_{di}}{\int_{-\infty}^{\infty} f(F_{di})\, f(\mathbf{y}_i|\mathbf{z}_i, F_{di})\, \mathrm{d}\, F_{di}} \\ &\approx \frac{\sum_{b_d=1}^{B} F_{b_d}\ f(\mathbf{y}_i|\mathbf{z}_i, F_{b_d})\, P_{b_d}}{\sum_{b_d=1}^{B}\ f(\mathbf{y}_i|\mathbf{z}_i, F_{b_d})\, P_{b_d}}.\end{aligned}$$

In Markov and multilevel models with covariates, the *Model Classification* output section reports the State and GClass membership probabilities given time-varying and group-level covariates, $\widehat{P}(x^d|\mathbf{z}_{it})$ and $\widehat{P}(x^g|\mathbf{z}^g_j)$. In Markov models, one also obtains joint classification: that is, for State given Class and State given previous State and Class.

14.11 Output-to-file Options

The *Standard Classification* option can be used to write the CFactors and GCFactors posterior means, the GClasses posterior probabilities, and the modal GClass to an output file. In models with GClasses, *Model Classification* saves the classification of groups into GClasses based on group covariates to the output file. In Markov models, one obtains classification information for Classes and States, as well as joint classification, which is State given Class and State given previous State and Class.

The *Individual Coefficients* corresponding to CFactor effects are computed in a special way:

$$\widehat{\lambda}_{iqd} = \sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i) \left[\widehat{E}(F_{di}|\mathbf{z}_i, \mathbf{y}_i, x)\, \widehat{\lambda}_{xqd}\right],$$

where $\widehat{E}(F_{di}|\mathbf{z}_i, \mathbf{y}_i, x)$ is the posterior mean of F_{di} given that i belongs to latent class x. The $\widehat{\lambda}_{iqd}$ can be used together with the $\widehat{\beta}_{iq}$ to obtain HB-like predicted values for case i. The posterior standard deviation of $\widehat{\lambda}_{iqd}$ equals

$$\widehat{\sigma}_{\widehat{\lambda}_{iqd}} = \sqrt{\sum_{x=1}^{K} \widehat{P}(x|\mathbf{z}_i, \mathbf{y}_i) \left[\widehat{E}(F_{di}|\mathbf{z}_i, \mathbf{y}_i, x)\, \widehat{\lambda}_{xqd} - \widehat{\lambda}_{iqd}\right]^2},$$

HB-like individual coefficients for a "full" intercept or predictor term may also be obtained by summing the various individual coefficient components for that term. For example, for a random-intercept model such as given in equation (35), the HB-like individual coefficient for the "full" intercept is computed by summing $\widehat{\beta}_{i0}$ and $\widehat{\lambda}_{i01}$.

In multilevel models, the *Cook's D* value is computed per group rather than per case. Thus, rather than for detecting influential cases, it can be used for detecting influential groups.

Part III: Syntax Model and Output Options

15 Introduction to Part III: Syntax Models

With the Latent GOLD Syntax module, you can specify many types of additional models compared to the Basic and Advanced modules. The most important additional modeling and output options are:

- The flexibility through the intuitive LG-Equations language which allows among others defining conditional effects and interaction effects, and specifying starting values and parameter restrictions, such as fixed value and monotonicity constraints, and equality and ratio constraints across parameters.

- A syntax model may contain any combination of nominal, ordinal, and continuous latent variables. This expands upon the GUI modules, which allow the use of either a single nominal latent variable (Cluster/Regression/Choice/Markov) or multiple ordinal latent variables (DFactor), possibly combined with continuous factors (CFactors). With the syntax module, one can specify easily, for example, a model with two nominal latent variables or with one nominal and one ordinal latent variable. Another example is a model with ordinal latent variables at the group level.

- The modeling of continuous latent variables is more general than what is possible in the GUI models. In LG-Syntax models, continuous latent variables can have free variances and be correlated. Also, it is possible to regress continuous latent variables on independent variables, discrete latent variables, and higher-level continuous latent variables. This makes it possible to specify factor analysis and item response models with covariates, as well as factor mixture models in which factor means and (co)variances differ across latent classes, and many other kinds of models.

- Discrete (nominal/ordinal) latent variables can be used in a path model in which one discrete latent variable is used as a predictor in the regres-

sion model for another discrete latent variable. Discrete latent variables may also be predicted by continuous latent variables.

- For models with multiple observations per case (models with a case id and repeated measurements for each id) it is possible to specify more than one dependent variable. This, in fact, expands the Regression and Choice modules by allowing for multiple dependent variables. This feature may be used to specify growth models for multivariate responses, as well as other kinds of multivariate regression models. An application in choice modeling is the simultaneous modeling of choice variable with other types of dependent variables (e.g., ratings or counts).

- Latent Markov models, also known as the hidden Markov or latent transition models, can be specified with the Markov GUI submodule, which is part of Latent GOLD 5.1 Advanced. LG-Syntax 5.1 provides various more extended models with dynamic latent variables, such as latent Markov models with multiple dynamic latent variables, second-order models, and multilevel variants of the latent Markov model (see, for example, Crayen et al. 2012). The dynamic latent variable option can also be used to define multilevel LC models with arbitrary numbers of levels.

- Another modeling option in LG Syntax 5.1 is the possibility to define continuous-time latent Markov models. These are latent Markov models that accommodate measurements which are not equally spaced in time.

- Latent GOLD 5.1 Basic includes a Step3 GUI submodule for bias adjusted step-three modeling (and scoring). LG-Syntax 5.1 provides various more extended step-three modeling possibilities, such a step-three models with both covariates and dependent variables and/or multiple latent variables, step-three latent Markov models, and step-three multilevel LC models. Moreover, the default rescaling of the posteriors can be suppressed by using the keyword "norescale" which may be useful if one specifies a step3 model using only a subset of the original latent classes.

- Three additional starting values options are the possibility to use Newton iterations (in addition to EM) in the starting values procedure; to

use starting values for factor loading based on a Principal Components Analysis (PCA) of the full sample, and to use an annealing EM algorithm with more fuzzy posteriors during the first 100 EM iterations of the start sets and where fuzzyness is reduced every 5 iterations.

- For ordinal dependent variables, in additional to the adjacent-category logit model, in LG-Syntax one can specify models based on cumulative responses probabilities – cumulative logit, probit, log-log, linear and log-linear models – as well as models for continuation-ratio or sequential logits. The probit, log-log, linear, and log-linear specifications also yield new models for dichotomous responses.

- LG-Syntax 5.1 contains various additional modeling options for continuous dependent variables. In addition to linear regression models with normal error distributions, one can also use gamma, beta, or von Mises distributions, for non-negative, 0-1 range, and circular variables, respectively.

- In LG Syntax 5.1 is the possibility to specify a log-linear scale factor model for categorical response variables (nominal, all ordinal types, all choice types). Since the scale factor is inversely related to the response uncertainty, this option makes it possible to model heterogeneity in response (un)certainty.

- There are various options to influence the coding and/or meaning of parameter sets. These are referred to as ~nom, ~ord, ~squ, ~dif, ~ful, ~tra, ~err, ~wei, ~int, and ~mis. The option ~wei, for example, allows defining models with a cell weight vector.

- In LG-Syntax, it is possible to specify user-defined Wald tests, to perform (asymptotic) power computation for Wald and likelihood-ratio tests, and to request Score tests and EPCs (expected parameter changes) for restricted parameters.

- LG-Syntax implements a rather general method for standard error computation in stepwise estimation; that is, to account for the additional variance caused by using estimates from a previous step as fixed parameters (Gong and Samaniego 1981; Oberski and Satorra, 2013). This method was used by Bakk, Oberski, and Vermunt (2014) for the adjustment of standard errors in three-step latent class analysis by taking

into account that the third step uses quantities based on parameters estimated in the first step, which themselves have a particular variance-covariance matrix. For more information on this approach, consult the above references and the LG-Syntax guide.

- In Latent GOLD GUI models, the label switching problem is resolved by reporting latent classes sorted from large to small, respecting the imposed constraints. With the LG-Syntax output option "reorder", one achieves the same.

- There are three additional Bayes options available in LG-Syntax: per-class, percategory, and uniform (see section on Bayes constants or the LG-Syntax manual for more information).

- LG-Syntax implements a procedure to obtain goodness-of-fit tests under the MAR (missing at random) assumption which also underlies maximum likelihood estimation of the model parameters.

Latent GOLD Syntax not only contains additional modeling and output options, it also implements options for Monte Carlo simulation, multiple imputation,[54] n-fold cross-validation, writing detailed output to text files, and re-using models with saved parameters. The Syntax Guide provides a complete description of Syntax and explains how to use the LG-Syntax system.

The remainder of Part III of the Technical Guide provides technical information on various of the modeling and output options which are specific to the Latent GOLD Syntax system. The next section discusses alternative regression models for dichotomous, ordinal, and continuous variables, log-linear scale factor models, regression models with cells weights, and continuous-time latent Markov models. Then, attention is paid to alternative complex sampling variance estimators, power computation for Wald and likelihood-ratio tests, Score tests and EPCs, adjusting chi-squared tests for missing values, and identification checking.

[54] Various multiple imputation methods using a LC model are implemented. These include bootstrapping, divisive LC analysis, Gibbs sampling, and using a Dirichlet Process prior (Vidotto, Kaptein, Vermunt, 2015).

16 Various Modeling Options which are Specific to Syntax

This section provides technical information on various modeling options which are specific to the LG-Syntax:

1. Alternative regression models for dichotomous and ordinal dependent variables;
2. Alternative regression models for continuous dependent variables;
3. Log-linear scale factor models for categorical dependent variables;
4. Regression models with cell weights (w̃ei option);
5. Continuous-time Markov models.

16.1 Alternative Regression Models for Dichotomous and Ordinal Dependent Variables

In Latent GOLD Basic and Advanced modules, a dichotomous dependent variable is modeled using a binary logit model, by defining it to be either nominal, ordinal, or (when 0/1 coded) a binomial count. An ordinal dependent variable is modeled in the GUI using an adjacent-category ordinal logit model, which is a restricted multinomial logit model (see section 2.3). LG syntax implements six alternative regression models for dichotomous and ordinal response variables:

1. the cumulative logit model (clogit),
2. the cumulative probit model (probit),
3. the cumulative negative log-log model (loglog1),
4. the cumulative complementary log-log model (loglog2),
5. the cumulative linear model (lin)
6. the cumulative log-linear model (log)
7. the continuation-ratio or sequential logit model (seqlogit1), and

8. a second variant of the continuation-ratio or sequential logit model (seqlogit2).

While these models can be used for ordinal response variables, when responses are dichotomous, they reduce to the binary logit (1, 5, and 6), probit (2), and log-log models (3 and 4), respectively. Each of the four cumulative models uses a cumulative link function transforming the probability of responding in category m or higher, $P(y \geq m|\mathbf{z}_i)$, into the linear predictor, $\eta_{m|\mathbf{z}_i}$. Using the notation from section 2.3,

$$\begin{aligned}
\eta_{m|\mathbf{z}_i} &= \log \frac{P(y \geq m|\mathbf{z}_i)}{P(y < m|\mathbf{z}_i)}, \\
\eta_{m|\mathbf{z}_i} &= \text{InverseCumulativeNormal}[P(y \geq m|\mathbf{z}_i)], \\
\eta_{m|\mathbf{z}_i} &= -\log\{-\log[P(y \geq m|\mathbf{z}_i)]\}, \\
\eta_{m|\mathbf{z}_i} &= \log\{-\log[1 - P(y \geq m|\mathbf{z}_i)]\}, \\
\eta_{m|\mathbf{z}_i} &= P(y \geq m|\mathbf{z}_i), \text{ and} \\
\eta_{m|\mathbf{z}_i} &= \log[P(y \geq m|\mathbf{z}_i)],
\end{aligned}$$

are the link functions for the cumulative logit, probit, negative log-log, complementary log-log, linear, and log-linear models, respectively.

The corresponding inverse link functions which transform $\eta_{m|\mathbf{z}_i}$ back to $P(y \geq m|\mathbf{z}_i)$ are

$$\begin{aligned}
P(y \geq m|\mathbf{z}_i) &= \frac{\exp(\eta_{m|\mathbf{z}_i})}{1 + \exp(\eta_{m|\mathbf{z}_i})}, \\
P(y \geq m|\mathbf{z}_i) &= \text{CumulativeNormal}(\eta_{m|\mathbf{z}_i}), \\
P(y \geq m|\mathbf{z}_i) &= \exp(-\exp(-\eta_{m|\mathbf{z}_i})], \\
P(y \geq m|\mathbf{z}_i) &= 1 - \exp[-\exp(\eta_{m|\mathbf{z}_i})], \\
P(y \geq m|\mathbf{z}_i) &= \eta_{m|\mathbf{z}_i}, \text{ and} \\
P(y \geq m|\mathbf{z}_i) &= \exp(\eta_{m|\mathbf{z}_i}),
\end{aligned}$$

respectively. The linear model for $\eta_{m|\mathbf{z}_i}$ has the following form:

$$\eta_{m|\mathbf{z}_i} = \beta_{0m} + \sum_{p=1}^{P} \beta_p z_{pi}.$$

As in the adjacent-category ordinal logit model, intercepts (called thresholds) are category specific, whereas slopes are constant across categories. The β_p

parameters have the expected interpretation: the higher the values of z_{pi} the more likely the higher responses. It is important to note that other software may instead model $P(y \leq m|\mathbf{z}_i)$, sometimes with reversed signs for the slope parameters. Whereas our parameterization yields decreasing thresholds, the other parameterization yields increasing thresholds, but the two are equivalent except for the sign. Another difference is that our negative log-log corresponds to a complementary log-log for $P(y \leq m|\mathbf{z}_i)$ and vice versa. Note that the negative log-log model is equivalent to a complementary log-log model for a reverse coded dependent variable.

Note that the identity link (the linear model) does not guarantee that probabilities stay in the 0-1 range and the log link does not guarantee that probabilities are smaller than 1. Therefore, special care should be taken when specifying (user-defined) starting values. Moreover, there is no guarantee that the algorithms will converge. On the other hand, sometimes it is convenient to define constraints on probabilities in linear or log-linear form, which is why we implemented these models.

The sequential logit model (seqlogit1) is a model for

$$\eta_{m|\mathbf{z}_i} = \log \frac{P(y \geq m|\mathbf{z}_i)}{P(y = m-1|\mathbf{z}_i)},$$

or

$$P(y \geq m|\mathbf{z}_i; y \geq m-1) = \frac{\exp(\eta_{m|\mathbf{z}_i})}{1+\exp(\eta_{m|\mathbf{z}_i})},$$

that is, for the probability of answering category m or higher conditional on answering category m-1 or higher. The second type of sequential logit model (seqlogit2) is a model for

$$\eta_{m|\mathbf{z}_i} = \log \frac{P(y = m|\mathbf{z}_i)}{P(y < m|\mathbf{z}_i)},$$

or

$$P(y = m|\mathbf{z}_i; y \leq m) = \frac{\exp(\eta_{m|\mathbf{z}_i})}{1+\exp(\eta_{m|\mathbf{z}_i})},$$

that is, for the probability of answering category m conditional on answering category m or lower. It should be noted that second model is equivalent to the first one after reversing the order of the categories of the dependent variable.

For more information on the ordinal regression models we refer to textbooks on categorical data analysis. More specifically, see chapter 7 of Agresti's (2002) excellent "Categorical Data Analysis" book.

16.2 Additional Regression Models for Continuous Dependent Variables

Three alternative regression models are implemented for specific types of continuous dependent variables. These are:

1. A gamma regression model for continuous variables taking on non-negative values. The link function is log, meaning that the expected value $\mu_{t,x,\mathbf{z}_i}$ is modeled with a linear model after a log transformation.

2. A beta regression model for continuous variables taking on values between 0 and 1 (excluding these numbers). The regression model for $\mu_{t,x,\mathbf{z}_i}$ uses a logit link.

3. A von Mises regression model for circular data (data indicating the position on a circle) and ranging between 0 and 2π. The regression model for $\mu_{t,x,\mathbf{z}_i}$ uses an identity link.

Note that these are all two parameter distributions, implying that in addition to a regression equation for the expected values $\mu_{t,x,\mathbf{z}_i}$, there is a variance or dispersion parameter $\sigma^2_{t,x}$. To supplement section 2.2, we now describe the conditional distribution of y_{it} for the three types of continuous dependent variables in terms of $\mu_{t,x,\mathbf{z}_i}$ and $\sigma^2_{t,x}$.

The gamma distribution with mean $\mu_{t,x,\mathbf{z}_i}$, shape $\frac{1}{\sigma^2_{t,x}}$, and scale $\mu_{t,x,\mathbf{z}_i}\sigma^2_{t,x}$ is defined as follows:

$$f(y_{it}|x,\mathbf{z}_i) = \frac{1}{\Gamma\left(\frac{1}{\sigma^2_{t,x}}\right)\left(\mu_{t,x,\mathbf{z}_i}\sigma^2_{t,x}\right)^{\frac{1}{\sigma^2_{t,x}}}} \exp\left(-\frac{y_{it}}{\mu_{t,x,\mathbf{z}_i}\sigma^2_{t,x}}\right) y_{it}^{\frac{1}{\sigma^2_{t,x}}}.$$

Note that similar to models for overdispersed counts, we model the inverse of the scale rather than the scale itself.

The beta distribution with mean $\mu_{t,x,\mathbf{z}_i}$ and shape parameters $\frac{\mu_{t,x,\mathbf{z}_i}}{\sigma^2_{t,x}}$ and $\frac{1-\mu_{t,x,\mathbf{z}_i}}{\sigma^2_{t,x}}$ is defined as follows:

$$f(y_{it}|x,\mathbf{z}_i) = \frac{\Gamma\left(\frac{1}{\sigma^2_{t,x}}\right)}{\Gamma\left(\frac{\mu_{t,x,\mathbf{z}_i}}{\sigma^2_{t,x}}\right)\Gamma\left(\frac{1-\mu_{t,x,\mathbf{z}_i}}{\sigma^2_{t,x}}\right)} y_{it}^{\frac{\mu_{t,x,\mathbf{z}_i}}{\sigma^2_{t,x}}-1}(1-y_{it})^{\frac{1-\mu_{t,x,\mathbf{z}_i}}{\sigma^2_{t,x}}-1}.$$

The von Mises distribution with mean $\mu_{t,x,\mathbf{z}_i}$ and scale $\frac{1}{\sigma^2_{t,x}}$ has the following form

$$f(y_{it}|x,\mathbf{z}_i) = \frac{\cos(y_{it} - \mu_{t,x,\mathbf{z}_i})}{2\pi I_0\left(\frac{1}{\sigma^2_{t,x}}\right)\sigma^2_{t,x}},$$

where $I_0(\cdot)$ is the modified Bessel function of order 0.

16.3 Log-linear Scale Factor Models for Categorical Dependent Variables

An interesting modeling feature is the possibility to include a scale factor in regression models for categorical response variables; that is, in models for nominal and the various types of ordinal and choice dependent variables. A scale factor is a term by which all parameters in the regression model are multiplied, and which thus allows modeling proportionality of parameter values across groups. The inverse of the scale factor is proportional to the standard deviation of the error distribution when using the underlying latent variable interpretation of the categorical response regression model concerned. Therefore, this option makes it possible to model heterogeneity in response (un)certainty.

When the scale model is used the linear term $\eta_{m|\mathbf{z}_i}$ in the regression model for the categorical dependent variable is replaced by $\eta_{m|\mathbf{z}_i}\varphi_{\mathbf{z}_i}$, where $\varphi_{\mathbf{z}_i}$ represents the multiplicative scale factor. For example, in the case of a nominal dependent variable, the equation for the response probability for category m becomes

$$P(y_i = m|\mathbf{z}_i) = \frac{\exp(\eta_{m|\mathbf{z}_i}\varphi_{\mathbf{z}_i})}{\sum_{m'=1}^{M}\exp(\eta_{m'|\mathbf{z}_i}\varphi_{\mathbf{z}_i})}.$$

The scale factor is modeled by a log-linear equation, yielding a flexible approach for modeling its dependence on latent and/or independent variables. Moreover, the log-linear equation guarantees that the scale factor remains non-negative. This implies that

$$\log\varphi_{\mathbf{z}_i} = \xi_0 + \sum_{p=1}^{P}\xi_p z_{pi};$$

or

$$\varphi_{\mathbf{z}_i} = \exp\left(\xi_0 + \sum_{p=1}^{P}\xi_p z_{pi}\right)$$

Vermunt (2013) provides a more detailed description of these log-linear scale models, and Magidson and Vermunt (2007), and Magidson, Thomas, and Vermunt (2009) present applications of latent class models with scale factors.

16.4 Regression Models with a Cell Weight Vector ("~wei" Option)

The specification "~wei" is used to include fixed parameter values in multiplicative form in models for nominal and ordinal latent or dependent variables. In the log-linear analysis literature, these are referred to as cell weights (Clogg and Eliason, 1987; Vermunt, 1997). A cell weight is equivalent to what is known as an offset in generalized linear modeling, but then in exponential form; thus, a cell weight equals exp(offset) or an offset equals log(cell weight) (Agresti, 2002). Among other uses, cell weights can be used to fix probabilities to specific values.

The model for a nominal or ordinal variable with a cell weight has the following form:

$$P(y_i = m|\mathbf{z}_i) = \frac{\exp(\eta_{m|\mathbf{z}_i})cw_{m|\mathbf{z}_i}}{\sum_{m'=1}^{M}\exp(\eta_{m'|\mathbf{z}_i})cw_{m'|\mathbf{z}_i}}.$$

The cell weight $cw_{m|\mathbf{z}_i}$ is a nonnegative term by which $\exp(\eta_{m|\mathbf{z}_i})$ – the exponent of the linear term – is multiplied. By default the cell weights equal 1, meaning that they have no impact. Note also that specifying a cell weight equal to 0 fixes the corresponding probability to 0.

We use the cells weights ourselves in some GUI models. That is, to define mover-stayer models, in which transition probabilities are fixed to 0 for the stayer class, to define manifest Markov models in which one indicator is perfectly related to the classes, and to define step-three models using ML adjustments in which the assigned class serves as an indicator with "known" response probabilities.

16.5 Continuous-Time Markov Models

LG-Syntax 5.1 contains an option for defining continuous-time (discrete-state) latent Markov models (Jackson and Sharples, 2002; Sharples et al., 2003; Böckenholt, 2005). In these models, the transition probabilities between measurement occasions are modeled as a function of transition intensity parameters and the length of the time interval. Transition intensities are

in fact equivalent to hazard rates in survival models; that is, the continuous-time equivalents of the probability of experiencing an event or transition (Cox and Miller, 1965; Kalbfleisch and Lawless, 1985).

When the data set contains observations at equidistant time intervals, one will typically use a discrete-time latent Markov model. However, in such situations, one can also use a continuous-time model, which will usually yield the same fit (same log-likelihood value), but the transition parameters take on a different interpretation. However, the continuous-time model is especially useful when observations do not occur at equidistant time intervals; that is, if the measurement occasions consist of snapshots of the underlying continuous-time process taken at arbitrary points in time. The only additional information that is needed to define a continuous-time latent Markov model in Latent GOLD 5.1 is information on the length of the time intervals between the measurement occasions.

We denote the transition intensity for the transition between origin state s and destination state r by $q_{r|s,\mathbf{z}_i}$ and the length of the time interval concerned by δ_{it}. The transition intensity matrix Q_i is a K by K matrix with the $q_{r|s,\mathbf{z}_i}$ as off-diagonal elements and the negative of the row sums of the off- diagonal elements on the diagonals. For K=3 (three latent states), $\mathbf{Q}_i$ equals

$$\mathbf{Q}_i = \begin{bmatrix} -q_{2|1,\mathbf{z}_i} - q_{3|1,\mathbf{z}_i} & q_{2|1,\mathbf{z}_i} & q_{3|1,\mathbf{z}_i} \\ q_{1|2,\mathbf{z}_i} & -q_{1|2,\mathbf{z}_i} - q_{3|2,\mathbf{z}_i} & q_{3|2,\mathbf{z}_i} \\ q_{1|3,\mathbf{z}_i} & q_{2|3,\mathbf{z}_i} & -q_{1|3,\mathbf{z}_i} - q_{2|3,\mathbf{z}_i} \end{bmatrix}.$$

Let$\mathbf{P}_{it}$ be a K by K matrix with the transition probabilities for person i at time interval t. The key relationship is the following:

$$\mathbf{P}_{it} = \mathbf{Exp}(\delta_{it}\mathbf{Q}_i),$$

where $\mathbf{Exp}(\cdot)$ is the matrix exponential. This equation shows the connection between the continuous-time transition intensities and the transition probabilities for the observed time intervals. As in discrete-time models, the transition probabilities define the model structure and the likelihood that is maximized to obtain the parameter estimates.

Because transition intensities are nonnegative, the $q_{r|s,\mathbf{z}_i}$ are modeled using a log-linear model; that is,

$$q_{r|s,\mathbf{z}_i} = \exp(\eta_{r|s\mathbf{z}_i}) = \exp(\gamma_{rs0} + \sum_{p=1}^{P} \gamma_{rsp} z_{pi}).$$

It should be noted that the γ parameters are defined only for $r \neq s$. Hence, whereas in discrete-time models, transition probabilities are functions of logit coefficients, in continuous-time models, they are functions of log transition intensity parameters and time interval lengths.

17 Various Output Options which are Specific to Syntax

This section provides technical information on various output options which are specific to the LG-Syntax:

1. Adjust chi-squared statistics with missing data;
2. Alternative variance estimators;
3. Power computation for Wald tests;
4. Power computation for likelihood-ratio tests;
5. Score tests and EPCs;
6. Identification checking.
7. Continuous-factor and random-effect covariances

17.1 Adjusted Chi-squared Statistics with Missing Data

In models for categorical dependent variables, Latent GOLD reports goodness-of-fit chi-squared statistics. However, with missing values on the dependent variables, these statistics provide a simultaneous test to both the model of interest and the assumption that the missing data is MCAR (missing completely at random) (Vermunt, 1997). LG-Syntax implements a simple procedure to obtain goodness-of-fit tests under the weaker MAR (missing at random) assumption which underlies maximum likelihood estimation of the model parameters.

The procedure begins by estimating the saturated model with the records with missing data included and computing the likelihood-ratio, Pearson, and Cressie-Read chi-squared statistics for this model. Since the saturated model itself fits perfectly, the resulting chi-squared statistics test only the MCAR assumption. By subtracting the resulting chi-squared values from those of the Latent GOLD model of interest and adjusting the number of degrees of freedom in the appropriate manner, we then obtain the corresponding chi-squared tests for the model of interest under MAR.

Chi-squared statistics adjusted for missing data are obtained with the output option "MARchi2". It should be noted that this option should not

be used in tables with more than 100,000 cells, because otherwise estimation of the saturated model may take very long. Moreover, chi-squared statistics for such tables will not be very meaningful.

These chi-squared adjustments can also be used when bootstrapping the distribution of the chi-squared statistics.

17.2 Other Variance Estimators

17.2.1 Complex sampling standard errors

The options for dealing with complex sampling designs are expanded in LG-Syntax beyond the capabilities of Latent GOLD Advanced. Specifically, in addition to obtaining complex sampling standard errors based on the Taylor linearization estimator (see section 13), LG-Syntax implements three alternative methods: jackknife, nonparametric bootstrap, and replicate weights. In these resampling procedures the elements of the variance-covariance matrix of the parameter estimates are obtained as follows:

$$\Sigma_{resampling,k,k'} = \sum_{r=1}^{R} \omega_r (\hat{\vartheta}_{kr} - \hat{\vartheta}_k)(\hat{\vartheta}_{k'r} - \hat{\vartheta}_{k'}),$$

where $\hat{\vartheta}_{kr}$ is the estimate of parameter k obtained with resample r, $\hat{\vartheta}_k$ is the maximum likelihood estimate, and ω_r is the weight given to resample r. The methods differ in the way the replicate samples are obtained and in the value of ω_r.

In the jackknife procedure, the replicates are obtained by leaving out one PSU (R is thus the number of PSUs in the sample) and upweighting the other PSUs in the same stratum by a factor $C_o/(C_o - 1)$, where C_o is the number of PSUs in stratum o. The value of ω_r is $(C_o - 1)/C_o$.

In a nonparametric bootstrap, replicate samples are obtained by sampling $C_o - 1$ PSUs with replacement from each stratum and the sampled PSUs are weighted by $C_o/(C_o - 1)$. The value of ω_r is $1/R$.

The last option is that of R sets of replicate weights, which are part of the survey. The R resamples are combined with a ω_r value equal to $1/(R{\cdot}(1-c)^2)$, where 1-c is the so-called average perturbation factor used in the construction of the weigths. We assume that this factor equals 0.5, yielding a ω_r equal to $4/R$.

It should be noted that although in theory this should not occur, when using the complex sampling standard error computation option, it may hap-

pen that some strata consist of a single PSU. In LG-Syntax implements three ways of dealing with this problem:

- certain: in the linearization estimator the single PSU has a contribution of 0 to the covariance of the gradients; in the jackknife and the bootstrap the single PSU is always included.
- scaled: is the same as certain but the variances are inflated by the fraction of certain strata.
- centered: in the linearization estimator gradients of the single PSU are compared with the overall gradients (which equal 0); in the jackknife the single PSU is treated in the same way as the other observations; in the bootstrap it is included with probability 0.5.

17.2.2 Other standard errors

The jackknife and nonparametric bootstrap can also be used without complex sampling. In such cases, there is a single stratum and the observations are the PSUs. In the jackknife procedure, the replicates will then thus be obtained by leaving out one observation, and in the nonparametric bootstrap by sampling observations with replacement from the total data set.

In LG-Syntax 5.1 it possible to obtain standard errors based on the expected information matrix. Note that this requires processing all possible data patterns, which is feasible only for frequency tables with not too many cells. In other situations, one may request the Monte Carlo version of the expected information matrix, which is obtained by simulating a large number of observations from the population defined by the model estimates. The expected information matrix is computed in a similar way as its approximate based on the outer-product of gradients (see equation 22), but involves a sum over all possible data patterns (I^*) and weighting by the estimated expected frequencies ($\widehat{m}_{i^*}$); that is, $\widehat{\Sigma}_{expected}(\boldsymbol{\vartheta}) = \widehat{\mathbf{B}}^{*-1}$, where element $B^*_{kk'}$ of $\mathbf{B}^*$ is defined as

$$B^*_{kk'} = \sum_{i^*=1}^{I^*} \widehat{m}_{i^*} \frac{\partial \log f(\mathbf{y}_{i^*}|\mathbf{z}_{i^*}, \boldsymbol{\vartheta})}{\partial \vartheta_k} \frac{\partial \log f(\mathbf{y}_{i^*}|\mathbf{z}_{i^*}, \boldsymbol{\vartheta})}{\partial \vartheta_{k'}}.$$

The Monte Carlo version $\mathbf{B}^{**}$ based on a sample of R observations (with

$R >> N$) is obtained as follows

$$B^{**}_{kk'} = \frac{N}{R}\sum_{r=1}^{R}\frac{\partial \log f(\mathbf{y}_r|\mathbf{z}_r,\boldsymbol{\vartheta})}{\partial \vartheta_k}\frac{\partial \log f(\mathbf{y}_r|\mathbf{z}_r,\boldsymbol{\vartheta})}{\partial \vartheta_{k'}}.$$

The expected information matrix is needed, among others, for the power computation for Wald statistics. Also Score tests and EPCs may be computed using the expected information matrix.

17.3 Power Computation for Wald Tests

In LG-Syntax 5.1, it possible to perform power computations for Wald tests, including for user-defined Wald test (Gudicha, Tekle, Vermunt, in press). Key in the computation of the power or the required sample size is the noncentrality parameter κ. It is defined as follows

$$\kappa = N\left(\mathbf{C}\boldsymbol{\vartheta}\right)'\left(\mathbf{C}\,\widehat{\Sigma}(\boldsymbol{\vartheta})\mathbf{C}'\right)^{-1}\left(\mathbf{C}\boldsymbol{\vartheta}\right),$$

where C defines the contrast of parameters ϑ to be tested by the Wald test and Σ is the parameter covariance matrix based on the expected information matrix (for a single observation). Let, moreover, CV_α be the critical value corresponding to type I error α obtained from the (central) chi-squared distribution with the appropriate number of degrees of freedom. The power 1-β given sample size N is the probably of obtaining a value larger than CV_α from the chi-squared distribution with non-centrality parameter κ.

To compute the required sample size N to achieve power 1-β, one determines the value of κ corresponding to 1-β, and computes N by

$$N = \frac{\kappa}{\left(\mathbf{C}\boldsymbol{\vartheta}\right)'\left(\mathbf{C}\,\widehat{\Sigma}(\boldsymbol{\vartheta})\mathbf{C}'\right)^{-1}\left(\mathbf{C}\boldsymbol{\vartheta}\right)}.$$

17.4 Asymptotic Power Computation for Likelihood-Ratio Tests

LG-Syntax can be used for computing the asymptotic power of likelihood-ratio tests using the procedure described in Gudicha, Schmittmann, and Vermunt (in press; under review). This involves defining and running H_1

and H_0 models using either an exemplary data set or a large simulated data set under the H_1 model.[55]

As in the power computation for Wald tests, we need to obtain the non-centrality parameter κ given sample size N. It is defined as follows:

$$\kappa = N\ 2[E(\log \mathcal{L}_1) - E(\log \mathcal{L}_0)],$$

where $E(\log \mathcal{L}_1)$ and $E(\log \mathcal{L}_0)$ represent the expected log-likelihood values (scaled to 1 observation) under the H_1 and H_0 model, respectively. The power 1-β is the probably of obtaining a value larger than CV_α from the chi-squared distribution with non-centrality parameter κ. Note that CV_α is the critical value corresponding to type I error α obtained from the (central) chi-squared distribution with the appropriate number of degrees of freedom.

For the computation of the sample size, we reverse the above relationship:

$$N = \frac{\kappa}{2[E(\log \mathcal{L}_1) - E(\log \mathcal{L}_0)]}$$

where κ is now the value of the noncentrality parameter corresponding with the required power level.

17.5 Score Tests and EPCs

LG-Syntax 5.1 implements a powerful tool for testing parameter constraints; that is, it reports a score test for each restriction and the expected parameter changes when a restriction is removed. This output can be requested with the output command "ScoreTest". In that case, one obtains:

1. A score test in the Parameters output for each restricted parameter set (Oberski, van Kollenburg, and Vermunt, 2013). Score tests are also called Lagrange multiplier tests or modification indices. Note that a score test is the estimated decrease of minus twice the log-likelihood value when relaxing the constraint of interest.

2. An EPC(self) value for each restricted parameter in the Parameters output (Oberski and Vermunt, under review). By EPC(self) we refer to the expected changes in the parameters of the restricted set itself after removing the restriction concerned.

[55] Note that it also possible to compute the power of likelihood-ratio tests using the Monte Carlo power option.

3. An output section nested within Parameters containing the EPC(other) values for all free parameters (Oberski, 2014; Oberski and Vermunt, 2013, Oberski, Vermunt, and Moors, 2015). These are the expected parameter changes in the other parameters when the restriction concerned is removed. EPC(other) statistics can be used for sensitivity analyses; that is, to determine whether a restriction affects the parameters of primary interest. We therefore also refer to these statistics as EPC(interest).

Let $\mathbf{\Sigma}_{11}$ denote the covariance matrix of the free (other) parameters. What we need in the computation of Score test and EPC's is the negative of second derivatives for the restricted (self) parameter set, $\mathbf{H}_{22}$, the negative of the cross-derivatives, $\mathbf{H}_{12}$, and the gradients of the restricted parameters $\mathbf{g}_2$. The Score test is obtained as

$$\mathbf{Score} = \mathbf{g}_2^{'}\mathbf{\Sigma}_{22}\mathbf{g}_2,$$

with

$$\mathbf{\Sigma}_{22} = \left(-\mathbf{H}_{22} - \mathbf{H}_{12}\mathbf{\Sigma}_{11}\mathbf{H}_{12}^{'}\right)^{-1}.$$

EPC(self) is defined as

$$\mathbf{EPC}(\text{self}) = \mathbf{\Sigma}_{22}\mathbf{g}_2,$$

and the **EPC**(other) equals

$$\mathbf{EPC}(\text{other}) = \mathbf{\Sigma}_{11}\mathbf{H}_{12}^{'}\mathbf{\Sigma}_{22}\mathbf{g}_2 = \mathbf{\Sigma}_{11}\mathbf{H}_{12}^{'}\mathbf{EPC}(\text{self}).$$

Score tests and EPCs can be requested in combination with various types of variance estimators (standard, robust, fast, and expected). For more information on this, see Oberski and Vermunt (under review).

17.6 Identification Checking

Certain latent class and mixture models for categorical dependent variables and/or binomial counts may not be identified. This means that the solution is not unique in the sense that different combinations of parameter values yield the same log-likelihood value, or equivalently, the same estimated probabilities. LG-Syntax contains a procedure to detect whether there are

such identification problems. This procedure is activated by including the command 'identification[=%d]' in the list of the output options.

The identification check implemented in LG-Syntax makes use of the Jacobian matrix (**J**). A model is *locally* identified if the rank of the Jacobian matrix equals the number of free model parameters. Conversely, if the rank is lower than the number of free model parameters, the model is not identified. Note that we used the term *local* identification, which refers to the fact that conditional on a particular set of parameters there is no other set in the direct neighborhood that gives the same model probabilities.

Element (i, k) of the Jacobian matrix is defined as

$$J_{ik} = \frac{\partial P(\mathbf{y}_i|\mathbf{z}_i)}{\partial \vartheta_k},$$

which is the first derivative of the probability of having data pattern i (belonging to cell i of the frequency table) towards parameter k. These derivatives are obtained for all possible data patterns (and if there are very many, for a large enough number of data patterns).

LG-Syntax computes the Jacobian and evaluates its rank multiple times (the default is 10 times) using different random parameter values. The reported 'number of non-identified parameters' is the rank deficiency that is encountered most of the time. The 'Iterationdetails' output listing reports the rank deficiency for each trial, from which it can be seen that typically most or all trials give the same rank deficiencies.

The rank of **J** is determined as follows: the square matrix $\mathbf{J}'\mathbf{J}$ is computed, the eigenvalues of this matrix are calculated, and number of eigenvalues which are at least 1.0e-15 times as large as the largest eigenvalue is counted. That is, eigenvalues less than 1.0e-15 times the largest correspond to rank deficiencies.

It should be noted that this procedure does not prove global identification, but instead checks an important necessary condition for global identification, which is local identification for certain sets of parameters. In practice, in turns out that, except for computational precision issues, it does not make a difference which parameter values are used to determine the rank of the Jacobian. More specifically, if a model contains two non-identified parameters, the rank of the Jacobian will be 2 less than the number of parameters regardless of the parameter values used to evaluate the Jacobian.

17.7 Continuous-Factor and Random-Effect Covariances

In Syntax models, the (co)variances of continuous factors are free parameters. However, when numerical integration is used, the estimated parameters are not the (co)variances themselves, but the elements of the matrix obtained by a Cholesky decomposition of the (co)variance matrix. Let us denote this lower-diagonal matrix by $\mathbf{C}_F$. The covariance matrix of the continuous latent variables is obtained as $\mathbf{\Sigma}_F = \mathbf{C}_F \mathbf{C}'_F$. These (co)variances are reported in the Parameters output.

In Regression GUI models, Latent GOLD provides the (co)variances of the random effects when the model contains continuous factors. These are obtained as follows: $\mathbf{\Sigma}_\Psi = \mathbf{\Lambda} \mathbf{\Lambda}'$, where $\mathbf{\Lambda}$ contains the effects of the continuous factors on the responses (the loadings). This formula holds when the variances of the continuous latent variables are fixed to 1 and the covariances are fixed to 0; that is, when the covariance matrix $\mathbf{\Sigma}_F$ is an identity matrix $\mathbf{I}$. In the more general case with free continuous factor (co)variances, the covariance matrix of the random effects can be obtained as follows: $\mathbf{\Sigma}_\Psi = \mathbf{\Lambda} \mathbf{\Sigma}_F \mathbf{\Lambda}'$.

17.8 The LTB Method for Distal Outcomes

Lanza, Tran, and Bray (2013) proposed a method for investigating the relationship between class membership and distal outcomes, which is especially useful with continuous distal outcomes (referred to as the LTB method). It involves using the distal outcome as a (numeric) covariate in the latent class model of interest, and subsequently computing output similar to what is provided in the Profile output in Cluster models.

In Syntax models, this output is obtained using the output option "profile=LTB". It yields the class-specific means and variances for the distal outcome(s) included in the model as (numeric) covariate(s). However, standard errors and tests provided are somewhat ad hoc (see, Bakk, Oberski, and Vermunt, in press). Bakk, Oberski, and Vermunt (in press) show that correct standard errors and tests can be obtained by combining the LTB approach with bootstrap standard errors, which is achieved with the output option standarderrors=npbootstrap.

The LTB procedure can also be combined with a three-step latent class analysis. This involves using the distal outcomes as covariates in the step3 model (Bakk, Oberski, and Vermunt, in press).

18 Bibliography

Agresti, A. (2002). *Categorical Data Analysis.* Second Edition, New York: Wiley.

Agresti A., Booth, J.G., Hobert, J.P., and Caffo, B. (2000). Random-effects modeling of categorical response data. *Sociological Methodology*, 30, 27-80.

Aitkin, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6, 251-262

Aitkin (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55, 218-234.

Andrews, R.L., and Currim, I.S. (2003). A Comparison of segment retention criteria for finite mixture logit models, *Journal of Marketing Research*, 40, 235-243.

Andrich, D. (1978): A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

Bakk, Z., Oberski, D.L., and Vermunt, J.K. (2014). . *Political Analysis*, 22, 520-540.

Bakk, Zs, Oberski, D.L., and Vermunt, J.K. (in press). Relating latent class membership to continuous distal outcomes: improving the LTB approach and a modified three-step implementation. *Structural Equation Modeling.*

Bakk, Zs., Tekle, F.B., and Vermunt, J.K. (2013). Estimating the association between latent class membership and external variables using bias adjusted three-step approaches. *Sociological Methodology.*

Bakk, Zs, and Vermunt, J.K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling*, 23, 20-31.

Banfield, J.D., and Raftery, A.E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics*, 49, 803-821.

Bartholomew, D.J., and Knott, M. (1999). *Latent Variable Models and Factor Analysis.* London: Arnold.

Bartholomew, D.J., and Tzamourani, P. (1999). The goodness-of-fit of latent trait models in attitude measurement. *Sociological Methods and Research*, 27, 525-546.

Bartolucci, F., Farcomeni, A. and Pennoni, F. (2013), *Latent Markov models for longitudinal data.* Chapman and Hall/CRC press.

Bartolucci, F., Pennoni, F., and Francis, B. (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society, Series A*, 170, 151-132.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41, 164-171.

Birnbaum, A. (1968). Some latent traits models and their use in inferring an examinee's ability. F.M. Lord and M.R. Novick, *Statistical Theories of Mental Test Scores*, 397-479. Reading, MA: Addison-Wesley.

Blossfeld, H.P., and Rohwer, G. (1995). *Techniques of event history modeling.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Bock, D.R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* , 37, 29-51.

Bock, R.D., and Aikin, M. (1981). Marginal maximum likelihood estimation of item parameters. *Psychometrika*, 46, 443-459.

Böckenholt, U. (1993). A latent class regression approach for the analysis of recurrent choices. *British Journal of Mathematical and Statistical Psychology,* 46, 95-118.

Böckenholt, U. (2001). Mixed-effects analyses of rank-ordered data. *Psychometrika*, 66, 45-62.

Böckenholt, U. (2005). A latent Markov model for the analysis of longitudinal data collected in continuous time: States, durations, and transitions, *Psychological Methods*, 10, 65-83.

Böhning, D. (2000). *Computer-Assisted Analysis of Mixtures: Meta-analysis, Disease Mapping and Others.* London: Chapmann & Hall.

Bolck, A., M.A. Croon, and J.A. Hagenaars. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, 12, 3-27.

Burns, N., D.R. Kinder, S.J. Rosenstone, V. Sapiro, and the National Election Studies. NATIONAL ELECTION STUDIES, 2000: PRE-/POST- ELECTION STUDY [dataset id:2000.T]. Ann Arbor, MI: University of Michigan, Center for Political Studies [producer and distributor], 2001. Can be downloaded at http://www.umich.edu/ nes/

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician*, 36, 153-157.

Celeux, G., and Govaert, G. (1992). A classification EM algorithm for clustering and two stochastic versions. *Computational Statistics & Data Analysis*, 14, 315–332.

Clogg, C.C. (1981). New developments in latent structure analysis. D.J. Jackson and E.F. Borgotta (eds.), *Factor Analysis and Measurement in Sociological Research*, 215-246. Beverly Hills: Sage Publications.

Clogg, C.C. (1988). Latent class models for measuring. R. Langeheine and J.Rost (eds.), *Latent Trait and Latent Class Models*. New York, London: Plenum Press.

Clogg, C.C. (1995). Latent class models. G.Arminger, C.C.Clogg, and M.E.Sobel (eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, 311-359. New York: Plenum Press.

Clogg, C. C., and Eliason, S. R. (1987). Some common problems in log-linear analysis. *Sociological Methods Research*, 15, 4-44.

Clogg, C.C., Rubin, D.R., Schenker, N., Schultz, B., Weidman, L. (1991). Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logit regression. *Journal of the American Statistical Association*, 86, 68-78.

Collins, L.M., Fidler, P.F., Wugalter, S.E., and Long, L.D. (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research*, 28, 375-389.

Collins, L.M., and Lanza, S.T. (2010). *Latent class and latent transition analysis for the social, behavioral, and health sciences*. New York: Wiley.

Collins, L M. and Wugalter, S.E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27, 131-157.

Cox, D.R., and Miller, H.D. (1965). *The theory of stochastic processes.* London: Methuen.

Crayen, C., Eid, M., Lischetzke, T.; Courvoisier, D.S., and Vermunt, J.K. (2012). Exploring dynamics in mood regulation - mixture latent Markov modeling of ambulatory assessment data. *Psychosomatic Medicine*, 74, 366-376.

Croon, M.A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology*, 43, 171-192.

Croon, M.A. (2002). Ordering the classes. J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, 137-162. Cambridge University Press.

Dayton, C.M. (1998). Latent class scaling analysis. Thousand Oakes: Sage Publications.

Dayton, C.M., and Macready, G.B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83, 173-178.

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Ser. B.*, 39, 1-38.

Dias, J.G. (2004). *Finite Mixture Models: Review, Applications, and Computer-intensive Methods*. Phd. Dissertation. Research School Systems, Organisation and Management (SOM), Groningen of University, The Netherlands.

Dias, J.G., and Vermunt, J.K. (2007). Latent class modeling of website users' search patterns: Implications for online market segmentation. *Journal of Retailing and Consumer Services*, 14 , 359-368.

Dillon, W.R., and Kumar, A. (1994). Latent structure and other mixture models in marketing: A integertaive survey and overview. R.P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, 352-388, Cambridge: Blackwell Publishers.

Dolan, C.V., and Van der Maas, H.L.J. (1997). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, 63, 227-253.

Elliot, D. S., Huizinga, D. and Menard, S. (1989). *Multiple Problem Youth: Delinquence, Substance Use and Mental Health Problems*. New York: Springer-Verlag.

Everitt, B.S. (1988). A finite mixture model for the clustering of mixed-mode data. *Statistics and Probability Letters*, 6, 305-309.

Everitt, B.S., and Hand, D.J. (1981). *Finite Mixture Distributions.* London: Chapman and Hall.

Follman, D.A., and Lambert, D. (1989). Generalizing logistic regression by non-parametric mixing. *Journal of the American Statistical Association*, 84, 295-300.

Formann, A.K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association*, 87, 476-486.

Formann, A.K., and Kohlmann,T. (1998). Structural latent class models. Sociological Methods and Research, 26, 530-565.

Fox, J.-P., and Glas, G.A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. Psychometrika, 66, 269-286.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. New York, NY: Springer.

Galindo-Garre, F., and Vermunt, J.K, (2005). Testing log-linear models with inequality constraints: a comparison of asymptotic, bootstrap, and posterior predictive p values. *Statistica Neerlandica*, 59, 82-94

Galindo-Garre, F., Vermunt, J.K., and Croon M.A. (2002). Likelihood-ratio tests for order-restricted log-linear models: A comparison of asymptotic and bootstrap methods. *Metodología de las Ciencias del Comportamiento*, 4, 325-337.

Galindo-Garre, F., Vermunt, J.K., and W. Bergsma (2004). Bayesian posterior estimation of logit parameters with small samples. *Sociological Methods and Research*, 33, 88-117.

Gelfand, A.E., Hills, S.E. Racine-Poone, A, and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85, 972-985.

Gill, P.E., Murray, W., and Wright, M.H. (1981). *Practical Optimization*. London: Academic Press.

Glas, C.A.W. (1999). Modification indices for the 2PL and the nominal response model. *Psychometrika*, 64, 273-294.

Gelman, Andrew, Carlin, John B., Stern, Hal .S., and Robin, Donald B. (1995). *Bayesian Data Analysis*. London: Chapman & Hall.

Goldstein, H. (1995). *Multilevel Statistical Models*. New York: Halsted Press.

Gong, G., and Samaniego, F.J. (1981). Pseudo maximum likelihood estimation: Theory and applications. *The Annals of Statistics*, 9, 861869.

Goodman, L.A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable: Part I - A modified latent structure approach. *American Journal of Sociology*, 79, 1179-1259.

Goodman, L.A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.

Goodman, L.A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537-552.

Gudicha, D.W., Schmittmann, V.D., and Vermunt, J.K. (in press). Power computation for likelihood ratio tests for the transition parameters in latent Markov models. *Structural Equation Modeling*.

Gudicha, D.W., Schmittmann, V.D., and Vermunt, J.K. (under review). Statistical power of likelihood-ratio and Wald tests in latent class models with covariates.

Gudicha, D.W., Tekle, F.B., and Vermunt, J.K. (in press). Power and sample size determination for latent class models. *Journal of Classification*.

Gudicha, D.W., and Vermunt, J.K. (2013). Mixture model clustering with covariates using adjusted three-step approaches. B. Lausen, D. van den Poel, and A. Ultsch (eds), *Algorithms from and for Nature and Life; Studies in Classification, Data Analysis, and Knowledge Organization*. 87-93. Springer-Verlag GmbH, Heidelberg.

Haberman, S.J. (1979). *Analysis of Qualitative Data, Vol 2, New Developments*. New York: Academic Press.

Haberman, S.J. (1988). A stabilized Newton-Raphson algorithm for log-linear models for frequency tables derived by indirect observations. C.Clogg (ed.), *Sociological Methodology 1988*, 193-211. San Francisco: Jossey-Bass.

Hadgu, A., and Qu, Y. (1998). A biomedical application of latent class models with random effects. *Applied Statistics*, 47, 603-616.

Hagenaars, J.A. (1988). Latent structure models with direct effects between indicators: local dependence models. *Sociological Methods and Research*, 16, 379-405.

Hagenaars, J.A. (1990). *Categorical Longitudinal Data - Loglinear Analysis of Panel, Trend and Cohort Data*. Newbury Park: Sage.

Hagenaars, J.A. (1993). *Loglinear Models with Latent Variables*. Newbury Park: CA: Sage.

Hagenaars., J.A., and McCutcheon, A.L. (eds.) (2002). A*pplied Latent Class Analysis*, Cambridge: Cambridge University Press.

Hambleton, R.K., and Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Norwell, MA: Kluwer Academic Publishers.

Hamilton, J.D. (2008). Regime-switching models. In: S. Durlauf and L. Blume (eds.), *New Palgrave dictionary of economics, 2nd edition*. Palgrave McMillan Ltd.

Hedeker, D. (1998). *MIXPREG: a computer program for mixed-effect Poisson regression*. University of Illinois at Chicago.

Hedeker, D. (2003). A mixed-effects multinomial logistic regression model. *Statistics in Medicine*, 22, 1433-1446.

Hedeker, D. and Gibbons. R.D. (1996). MIXOR: A computer program for mixed effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine*, 49, 157-176.

Heinen, T. (1996). *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oakes: Sage Publications.

Huq, N. M., and Cleland, J. (1990). *Bangladesh Fertility Survey 1989 (Main Report)*. Dhaka: National Institute of Population Research and Training.

Hunt, L, and Jorgensen, M. (1999). Mixture model clustering using the MULTIMIX program. *Australian and New Zealand Journal of Statistics*, 41, 153-172.

Im, S., and Gionala, D. (1988). Mixed models for binomial data with an application to lamb mortality, *Applied Statistics*, 37, 196-204.

Imbens, G.W., and Rubin, D.B. (1997). Estimating outcome distributions for compliers in instrumental variable models, *Review of Economic Studies*, 64, 555-574.

Jackson C.H., and Sharples. L.D. (2002). Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine*, 21, 113-128. 2002.

Jedidi, K., Jagpal, H.S., and DeSarbo, W.S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39-59.

Kalbfleisch, J.D., and Lawless, J.F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80, 863-871.

Kamakura, W.A., Wedel, M., and Agrawal, J. (1994). Concomitant variable latent class models for the external analysis of choice data. *International Journal of Marketing Research*, 11, 541-464.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, 44, 457-467.

Koch, G.G., Landis, J.R., Freeman, J.L., Freeman, D.H., and Lehnen, R.G. (1977). A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics*, 33, 133-158.

Kolhmann T., and Formann A.K. (1997). Using latent class models to analyze response patterns in epidemiologic mail surveys. J.Rost and R.Langeheine (eds.). A*pplications of Latent Trait and Latent Class Models in the Social Sciences*. Muenster: Waxmann Verlag.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association*, 73, 805-811.

Land, C.K., McCall, P.L., and Nagin, D.S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models. *Sociological Methods and Research*, 24, 387-442.

Landis, J.R., and Koch, G.G. (1977). The measurement of observer agreement for categorical data, *Biometrics*, 33, 159?174.

Langeheine, R., Pannekoek, J., and Van de Pol, F. (1996). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*, 24, 492-516.

Langeheine, R., and Rost, J. (1988). *Latent Trait and Latent Class Models*. New York: Plenum Press.

Lanza, T.S., Tan, X., and Bray, C.B. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, 20, 1-26.

Lawrence C.J., Krzanowski, W.J. (1996). Mixture separation for mixed-mode data. *Statistics and Computing*, 6, 85-92.

Lazarsfeld, P.F., and Henry, N.W. (1968). Latent structure analysis. Boston: Houghton Mill.

Lenk, P.J., and DeSarbo, W.S. (2000). Bayesian inference for finite mixture models of generalized linear models with random effects, *Psychometrika*, 65, 93-119.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, 50, 325-335.

Little, R.J., and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oakes, CA: Sage Publication.

Luke, D.A. (2004). *Multilevel Modeling.* Thousand Oakes: Sage Publications

Lukociene, O., Varriale, R., and Vermunt, J.K.. (2010). The simultaneous decision(s) about the number of lower- and higher-level classes in multilevel latent class analysis. *Sociological Methodology*, 40, 247-283.

Lukociene, O., Vermunt, J.K (2010). Determining the number of components in mixture models for hierarchical data. In: Fink, A., Lausen, B., Seidel, W. and Ultsch, A. (eds.), *Advances in data analysis, data handling and business intelligence*, 241-249. Springer: Berlin-Heidelberg.

Lystig, T. C. and Hughes, J. (2002). Exact computation of the observed information matrix for hidden Markov models. *Journal of Computational and Graphical Statistics*, 11, 678-689.

MacDonald, I. L. and Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. Chapman and Hall, London

Masters, G. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Magidson, J. (1981). Qualitative variance, entropy, and correlation ratios for nominal dependent variables. *Social Sciences Research*, 10, 177-194.

Magidson, J. (1996). Maximum likelihood assessment of clinical trials based on an ordered categorical response. *Drug Information Journal,* Maple Glen, PA: Drug Information Association, Vol. 30, No. 1.

Magidson, J., Thomas, D., and Vermunt, J.K. (2009). A new model for the fusion of maxdiff scaling and ratings data. *Sawtooth Software Conference Proceedings*, *March 2009*, 83-103.

Magidson, J., and Vermunt, J.K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays, *Sociological Methodology*, 31, 223-264.

Magidson, J., and Vermunt, J.K. (2002a). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, 20, 37?44.

Magidson, J. and Vermunt, J.K. (2002b). Latent class modeling as a probabilistic extension of K-means clustering. *Quirk?s Marketing Research Review*, March 2002, 20 and 77-80.

Magidson, J., and Vermunt, J.K. (2003a). Comparing latent class factor analysis with the traditional approach in datamining. H. Bozdogan (ed.), *Statistical Data Mining and Knowledge Discovery,* 373-383. Boca Raton: Chapman & Hall/CRC.

Magidson, J., and Vermunt J.K. (2003b) A nontechnical introduction to latent class models. White Paper. Statistical Innovations.

Magidson, J., and Vermunt, J.K, (2004). Latent class analysis. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Chapter 10, 175-198. Thousand Oakes: Sage Publications.

Magidson, J., and Vermunt, J.K. (2005). An extension of the CHAID tree-based segmentation algorithm to multiple dependent variables. In: C. Weihs and

W. Gaul, *Classification: The Ubiquitous Challenge*, 176-183. Heidelberg: Springer.

Magidson, J., and Vermunt, J.K. (2007). Removing the scale factor confound in multinomial logit choice models to obtain better estimates of preference. *Sawtooth Software Conference Proceedings, October 2007*, 139-154.

Magidson, J., Vermunt, J.K., and Tran, B. (2009). Using a mixture latent Markov model to analyze longitudinal U.S. employment data involving measurement error. In: K. Shigemasu, A. Okada, T. Imaizumi, and T. Hoshino (eds.), *New Trends in Psychometrics,* 235-242. Universal Academy Press, Inc.

Mare, R.D. (1994). Discrete-time bivariate hazards with unobserved heterogeneity: a partially observed contingency table approach. P.V. Marsden (ed.), *Sociological Methodology 1994*, 341-385. Oxford: Basil Blackwell.

McCullagh, P., and Nelder, J.A. (1983). *Generalized Linear Models*. London: Chapman & Hall, second edition 1989.

McCutcheon, A.L. (1987). *Latent Class Analysis*, Sage University Paper. Newbury Park: Sage Publications.

McCutcheon, A.L. (1988). Sexual morality, pro-life values and attitudes toward abortion. *Sociological Methods and Research*, 16, 256-275.

McCutcheon, A.L. (1994). Latent logit models with polytomous effect variables. A. von Eye and C.C Clogg (eds.), *Latent Variable Analysis*. Newbury park, CA: Sage.

McFadden, D. and Train, D. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*,

McGrath, K., and Waterton, J. (1986). *British social attitudes, 1983?1986 panel survey* (Technical report). London: Social and Community Planning Research.

McLachlan, G.J., and Basford, K.E. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker.

McLachlan, G.J., and Krishnan, T. (1997). *The EM Algorithm and Extensions*. New York: John Wiley & Sons, Inc.

McLachlan, G.J., and Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley & Sons, Inc.

McKnight, B., and Van Den Eeden, S.K. (1993). A conditional analysis for two-treatment multiple period crossover designs with binomial or Poisson outcomes and subjects who drop out, *Statistics in Medicine*, 12, 825-834.

Mellenbergh, G.J. (1995). Conceptual notes on models for discrete polytomous item responses, *Applied Psychological Measurement* , 19, 91-100.

Meulders, M., De Boeck. P., Kuppens, P., and Van Mechelen, I. (2002). Constraint latent class analysis of three-way three-mode data, *Journal of Classification*, 19, 277-302.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *The British Journal of Mathematical and Statistical Psychology*, 49, 313-334.

Moustaki, I. and Knott, M. (2000). Generalized Latent Trait Models, *Psychometrika*, 65, 391-412.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16, 159-177.

Muthén, B. (2004). Latent variable analysis. Growth mixture modeling and related techniques for longitudinal data. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*, Chapter 19, 345-368. Thousand Oakes: Sage Publications.

Nagelkerke, E., Oberski, D.L., and Vermunt (in press). Goodness-of-fit measures for multilevel latent class models. *Sociological Methodology*.

Nylund, K.L., Muthen, M., and Muthen, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.

Oberski, D.L. (2014), Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models, *Political Analysis*, 22, 45–60.

Oberski, D.L, van Kollenburg, G.H., and Vermunt, J.K. (2013). A Monte Carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Classification and Data Analysis*, 7, 267-279.

Oberski, D. L., and Satorra, A. (2013). Measurement error models with uncertainty about the error variance. *Structural equation modeling*, 20, 409-428.

Oberski, D.L., and Vermunt, J.K. (under review). The expected parameter change (EPC) for local dependence assessment in binary data latent class models. Under review by *Psychometrika*.

Oberski, D.L., and Vermunt, J.K. (2013). A model-based approach to goodness-of-fit evaluation in item response theory. *Measurement*, 3, 117-122.

Oberski, D.L., Vermunt, J.K., and Moors, G.B.D. (2015). Evaluating measurement invariance in categorical data latent variable models with the EPC-interest, *Political Analysis*, 23, 550-563.

Paas, L. (2002). *Acquisition pattern analysis with Mokken scales: Applications in the financial services market*. Phd. Thesis. Tilburg University.

Paas, L. J., Vermunt, J. K., and Bijmolt, T. H. A. (2007). Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society, Series A*, 170, 955-974.

Palardy, G., and Vermunt, J.K.. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics*, 35, 532-565.

Patterson, B.H., Dayton,C.M.; Graubard, B.I. (2002). Latent class analysis of complex sample survey data: application to dietary data. *Journal of the American Statistical Association*, 97, 721-728.

Poulsen, C.A. (1982). *Latent structure analysis with choice modelling*. Aarhus: Aarhus School of Business Administratioon and Economics.

Poulsen, C.S. (1990). Mixed Markov and latent Markov modelling applied to brand choice behaviour. *International Journal of Research in Marketing*, 7, 519.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52, 797-810.

Rabe-Hesketh, S., Skrondal, A., and Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2, 1-21.

Ramos, S.B., Vermunt, J.K., and Dias, J.G. (2011). When markets fall down: Are emerging markets all equal? *International Journal of Finance and Economics*, 16, 324-338.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Data*. Chicago: The University of Chicago Press.

Reckase, M. (1997). A linear logistic multidimensional model for dichotomous response data. In W.J. van der Linden and R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory*, 271-286. New York: Springer.

Reiser, M. (1996). Analysis of residuals for the multinomial item response model. *Psychometrika*, 61, 509-528.

Reiser, M., and Lin, Y. (1999). Goodness-of-fit test for the latent class model when expected frequencies are small. M.Sobel and M.Becker (eds.), S*ociological Methodology 1999*, 81-111. Boston: Blackwell Publishers.

Rodriguez, G. and Goldman, N. (2001). Improved estimation procedures for multilevel models for binary response: a case study. *Journal of the Royal Statistical Society*, Series A, 164, 339-355.

Rost, J. (1990). Rasch models in latent classes. An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.

Rost, J., and Langeheine, R. (1997). *Applications of Latent Trait and Latent Class Models in the Social Sciences*. Muenster: Waxmann Verlag.

Schaeffer, N.C. (1988). An application of item response theory to the measurement of depression, Pp. 271-308 in *Sociological Methodology 1988*, edited by C. Clogg. Washington DC: American Sociological Association.

Schafer, J.L. (1997). *Statistical Analysis with Incomplete Data*. London: Chapman & Hall.

Sham, P. (1998). *Statistics in Human Genetics*. London: Arnold.

Sharples, L.D., Jackson, C.H., Parameshwar, J., Wallwork, J.,Large, S.R. (2003). Diagnostic accuracy of coronary angiography and risk factors for post-heart-transplant cardiac allograft vasculopathy. *Transplantation*, 76, 679–682.

Shockey, J. (1988). Latent class analysis: an introduction to discrete data models with unobserved variables. J.S. Long (ed.), *Common Problems in Quantitative Research*. Newbury Park, CA: Sage.

Simonoff, J.S. (2003). *Analyzing Categorical Data*. New York: Springer-Verlag

Skinner, C.J., Holt, D., and Smith, T.M.F. (eds.) (1989), *Analysis of Complex Surveys*, New York: Wiley.

Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models*. London: Chapman & Hall/CRC.

Smit, J.A., Kelderman, H., and Van der Flier, H. (2000). The mixed Birnbaum model: estimation using collateral information. *Methods of Psychological Research*, 5, 1-13.

Schmittmann, V.D., Dolan, C.V., van der Maas, H.L.J.,and Neale, M.C. (2005). Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research*, 40, 461-488.

Snijders, T.A.B., and Bosker, R.J. (1999). *Multilevel Analysis*. London: Sage Publications.

Steyer, R., and Partchev, I. (2001). Latent state-trait modeling with logistic item response models In: R. Cudeck, S. Du Toit, D. Sörbom (eds.), *Structural Equation Modeling: Present and Future*, 481-520. Chicago: Scientific Software International.

Stroud, A.H. & Secrest. D. (1966). Gaussian Quadrature Formulas. Englewood Cliffs, NJ: Prentice Hall.

Tay, L., Diener, E., Drasgow, F., and Vermunt, J.K. (2011). Multilevel mixed-measurement IRT analysis: An explication and application to self-reported emotions across the world. *Organizational Research Methods*, 14(1), 177-207.

Tay, L., Vermunt, J.K., and Wang, C. (2013). Assessing the item response theory with covariate (IRT-C) procedure for ascertaining differential item functioning. *International Journal of Testing*, 13, 201-222.

Thall, P. F, and Vail, S.C. (1990). Some covariance structure model for longitudinal count data with overdispersion. *Biometrics*, 46, 657-671.

Titterington, D.M., Smith, A.F., and Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. Chichester: John Wiley & Sons.

Uebersax, J.S. (1993). Statistical modeling of expert ratings on medical treatment appropriateness. *Journal of the American Statistical Association*, 88, 421-427.

Uebersax, J.S. (1997). Analysis of student problem behaviors with latent trait, latent class, and related probit mixture models. J.Rost and R.Langeheine (eds.). *Applications of Latent Trait and Latent Class Models in the Social Sciences.* Muenster: Waxmann Verlag.

Van der Ark, L.A., and Van der Heijden, P.G.M. (1998). Graphical display of latent budget and latent class analysis, with special reference to correspondence analysis. J. Blasius and M.Greenacre (eds.), *Visualization of Categorical Data*. Boston: Academic Press

Van de Pol, F., and De Leeuw, J. (1986). A latent Markov model to correct for measurement error. Sociological Methods and Research, 15, 118-141.

Van de Pol, F., and Langeheine, R. (1990) Mixed Markov latent class models. Sociological Methodology, 213-247.

Van der Heijden, P.G.M., Dessens, J., and Böckenholt, U. (1996). Estimating the concomitant-variable latent class model with the EM algorithm. *Journal of Educational and Behavioral Statistics*, 5, 215-229.

Van der Heijden P.G.M., Gilula, Z., and Van der Ark, L.A. (1999). On a relationship between joint correspondence analysis and latent class analysis. M.Sobel and M.Becker (eds.), *Sociological Methodology 1999*, 81-111. Boston: Blackwell Publishers.

Van der Heijden, P.G.M, Mooijaart, A., and De Leeuw, J. (1992). Constraint latent budget analysis. *Sociological Methodology*, 22, 279-320.

Van der Linden, W.J., and Hambleton, R.K. (eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer.

Van der Palm, D.W., van der Ark, L.A., and Vermunt, J.K. (in press a). A comparison of incomplete-data methods for categorical data. *Statistical Methods in Medical Research*.

Van der Palm, D.W., Van der Ark, L.A., and Vermunt, J.K. (in press b). Divisive latent class modeling as a density estimation method for categorical data. *Journal of Classification*.

Van Mierlo, H. (2003). *Self-managing Teams and Psychological Well-being*. Phd. dissertation. Eindhoven University of Technology, The Netherlands.

Varriale, R., and Vermunt, J.K. (2012). Multilevel mixture factor models. *Multivariate Behavioral Research*, 47, 247-275.

Vermunt, J.K. (1997). *Log-linear Models for Event Histories*. Thousand Oakes: Series QASS, vol 8. Sage Publications.

Vermunt, J.K. (1999). A general non-parametric approach to the analysis of ordinal categorical data. *Sociological Methodology*, 29, 197-221.

Vermunt, J.K. (2001). The use restricted latent class models for defining and testing nonparametric and parametric IRT models. *Applied Psychological Measurement*, 25, 283-294.

Vermunt, J.K.. (2002a). A general latent class approach for dealing with unobserved heterogeneity in the analysis of event history data. J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, 383-407. Cambridge: Cambridge University Press.

Vermunt, J.K. (2002b). Comments on Latent class analysis of complex sample survey data. *Journal of the American Statistical Association*, 97, 736-737.

Vermunt, J.K. (2002c). An Expectation-Maximization algorithm for generalised linear three-level models. *Multilevel Modelling Newsletter*, 14, 3-10.

Vermunt, J.K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213-239.

Vermunt, J.K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, 58, 220- 233.

Vermunt J.K. (2005). Mixed-effects logistic regression models for indirectly observed categorical outcome variables. *Multivariate Behavioral Research*, 40, 281-301.

Vermunt, J.K. (2007). Growth models for categorical response variables: standard, latent-class, and hybrid approaches. K. van Montfort, H. Oud, and A. Satorra (eds.). *Longitudinal Models in the Behavioral and Related Sciences.* 139-158. Mahwah, NJ: Erlbaum.

Vermunt, J.K. (2008a). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, 17, 33-51.

Vermunt, J.K. (2008b). Multilevel latent variable modeling: An application in educational testing. *Austrian Journal of Statistics*, 37 (3&4), 285-299.

Vermunt, J.K. (2010a). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469.

Vermunt, J.K (2010b). Longitudinal research using mixture models. In: K. van Montfort, J.H.L. Oud, and A. Satorra (eds.), *Longitudinal Research with Latent Variables*, 119-152. Heidelberg, Germany: Springer.

Vermunt, J.K.. (2011). K-means may perform as well as mixture model clustering but may also be much worse: Comment on Steinley and Brusco (2011). *Psychological Methods*, 16, 82-88.

Vermunt, J.K. (2013). Categorical response data. In: M.A. Scott, J.S. Simonoff, and B.D. Marx (Eds.), *The SAGE handbook of multilevel modeling*, 287-298. Thousand Oaks, CA: Sage.

Vermunt, J.K., and Hagenaars, J.A. (2004). Ordinal longitudinal data analysis. In: R. Hauspie, N. Cameron and L. Molinari (eds.). *Methods in Human Growth Research*, Chapter 15. Cambridge University Press.

Vermunt, J.K. Langeheine, R., and Böckenholt, U. (1999). Latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 179-207.

Vermunt, J.K., and Magidson, J. (2000a). *Latent GOLD 2.0 User's Guide.* Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., and Magidson. J. (2000b) Graphical displays for latent class cluster and latent class factor models. W. Jansen and J.G. Bethlehem (eds.),

Proceedings in Computational Statistics 2000, 121-122. Statistics Netherlands. ISSN 0253-018X.

Vermunt, J.K., and Magidson, J. (2001): *Latent Class Analysis with Sampling Weights*, Paper presented at the 6th annual meeting of the Methodology Section of the American Sociological Association, University of Minnesota, May 4-5, 2001.

Vermunt, J.K., and Magidson, J. (2002). Latent class cluster analysis. J.A. Hagenaars and A.L. McCutcheon (eds.), *Applied Latent Class Analysis*, 89-106. Cambridge: Cambridge University Press.

Vermunt, J.K. and Magidson, J. (2004). Latent class analysis. M. Lewis-Beck, A. Bryman, and T.F. Liao (eds.), *The Sage Encyclopedia of Social Science Research Methods*. NewBury Park: Sage Publications, Inc.

Vermunt, J.K. and Magidson, J. (2005a). Factor analysis with categorical indicators: a comparison between traditional and latent class approaches. In A. Van der Ark, M.A. Croon and K. Sijtsma (eds), *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*, 41-62, Mahwah: Erlbaum.

Vermunt, J.K, and Magidson, J. (2005b). Hierarchical mixture models for nested data structures. C. Weihs and W. Gaul (eds.), *Classification: The Ubiquitous Challenge*, 176-183. Heidelberg: Springer.

Vermunt, J.K. and Magidson, J. (2005c). Structural equation models: Mixture models. B. Everitt and D. Howell, (Eds.), *Encyclopedia of Statistics in the Behavioral Science*, 1922?1927. Wiley: Chichester, UK.

Vermunt, J.K., Rodrigo, M.F., Ato-Garcia, M. (2001) Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods and Research*, 30, 170- 196.

Vermunt, J.K., Tran, B., and Magidson, J. (2008). Latent class models in longitudinal research. In: S. Menard (ed.), *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, pp. 373-385. Burlington, MA: Elsevier.

Vermunt, J.K. and Van Dijk. L. (2001). A nonparametric random-coefficients approach: the latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.

Vermunt, J.K., Van Ginkel, J.R., Van der Ark, L.A., and, Sijtsma K. (2008). Multiple imputation of categorical data using latent class analysis. *Sociological Methodology*, 33, 369-297.

Vidotto, D., Kaptein, M.C., and Vermunt, J.K. (2015). Multiple imputation of missing categorical data using latent class models: State of art. *Psychological Test and Assessment Modeling*, 57, 542-576.

Visser, I. (2011). Seven things to remember about hidden markov models: A tutorial on Markovian models for time series. *Journal of Mathematical Psychology*, 55, 403-415.

Wedel, M., and DeSarbo, W.S (1994). A review of recent developments in latent class regression models. R.P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, 352-388, Cambridge: Blackwell Publishers.

Wedel, M., and Kamakura, W.A. (1998). *Market Segmentation: Concepts and Methodological Foundations*. Boston: Kluwer Academic Publishers.

Wedel, M., Kamakura, W.A., DeSarbo, W.S., and Ter Hofstede, F. (1995). Implications of asymmetry, nonproportionality, and heterogeneity in brand switching from piece-wise exponential mixture hazard models. *Journal of Marketing Research*, 32, 457-462.

Wedel, M., Ter Hofstede, F., and Steenkamp, J.-B.E.M. (1998). Mixture model analysis of complex samples, *Journal of Classification*, 15, 225-244.

Wolfe, J.H. (1970). Pattern clustering by multivariate cluster analysis. *Multivariate Behavioral Research*, 5, 329-350.

Wong, G.Y., and Mason, W.M. (1985). Hierarchical logistic models for multilevel analysis. *Journal of the American Statistical Association*, 80, 513-524.

Yamaguchi, K. (1991). *Event history analysis. Applied Social Research Methods*, Volume 28. Newbury Park: Sage Publications.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297-330.

Zackin, R., De Gruttola, V., and Laird, N. (1996). Nonparametric mixed-effects models for repeated binary data arising on serial dilution assays: An application to estimating viral burden in AIDS, *Journal of the American Statistical Association*, 91, 52-61.

Zwinderman, A.H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika*, 56, 589-600.

19 Notation

19.1 Basic Models

$f(\cdot)$	density
$P(\cdot)$	probability
i, I	case index, # of cases
t	indicator index or replication index
T	# of indicators
T_i	# of replications for case i
y_{it}	response of case i on indicator t or replication t
m, m_t	category of a nominal or ordinal response variable
y_m^*, y_m^{t*}	score assigned to category m of an ordinal response variable
h, H	index for a set of indicators, # of sets of indicators
T_h^*	# of indicators in set h
$\mathbf{y}_{ih}$	vector of responses in set h
x	nominal latent variable, a particular latent class
K	# of latent classes
ℓ, L	DFactor index, # of DFactors
x_ℓ	ordinal latent variable (DFactor), a particular level of DFactor ℓ
K_ℓ	# of level of DFactor x_ℓ
$x_{x_\ell}^{\ell*}$	score assigned to level x_ℓ of DFactor ℓ
r, R	covariate index, # of covariates
q, Q	predictor index, # of predictors
z_{ir}^{cov}	covariate
z_{itq}^{pred}	predictor
π_m	multinomial probability
π	binomial probability
θ	Poisson rate
μ	mean of continuous y_{it}
η	linear predictor
σ^2, $\mathbf{\Sigma}$	variance, variance-covariance matrix
β, γ	parameter in model for y_{it}, parameter in model for x or x_ℓ
τ_{ix}	known-class indicator
a_i	assigned class membership
o_i	outcome/dependent variable in step-three analysis
w_i	case weight
v_{it}	replications weight
u, U	"covariate" pattern index, # of "covariate" patterns
i^*, I^*	unique data pattern index, # of unique data patterns
N	total sample size (after weighting)

19.2 Advanced Models

x_t^d	latent State at time point t
K^d	number of latent States
t	time variable running from 0 to T_i
d, D	CFactor index, # of CFactors
F_{id}	scores of case i CFactor d
λ_d	an effect of CFactor d
j, J	group index, # of groups
I_j	# of cases in group j
y_{jit}	response of case i of group j on indicator (at replication) t
$\mathbf{y}_j$	vector of responses of group j
g	group-level quantity
x^g	group-level nominal latent variable, a particular group GClass
$z_{jr}^{g,cov}$	group-level covariate
F_{jd}^g	score of group j of group-level continuous factor (GCFactor) d
$\gamma^g, \beta^g, \lambda^g$	group-level parameters
o, O	stratum, # of strata
c, C_o	PSU, # of PSU's in stratum o
sw_{oci}	sampling weight
I_{oc}	# of cases in PSU c of stratum o
N_o	total # of PSUs in population in stratum o

19.3 Syntax Models

φ	scale factor
ϕ	log-linear scale factor parameter
cw	cell weight
δ	time interval
q	transition intensity
$\mathbf{Q}$	transition intensity matrix
$\mathbf{P}$	transition probability matrix
ω	weight of resample in variance formula
κ	non-centrality parameter
$\mathbf{J}$	Jacobian matrix